



**Europäisches
Patentamt**

**European
Patent Office**

**Office européen
des brevets**

RECEIVED

23 MAR 2004

WIPO

PCT

Bescheinigung

Certificate

Attestation

Die angehefteten Unterlagen stimmen mit der ursprünglich eingereichten Fassung der auf dem nächsten Blatt bezeichneten europäischen Patentanmeldung überein.

The attached documents are exact copies of the European patent application described on the following page, as originally filed.

Les documents fixés à cette attestation sont conformes à la version initialement déposée de la demande de brevet européen spécifiée à la page suivante.

Patentanmeldung Nr. Patent application No. Demande de brevet n°

03000249.7

PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH
RULE 17.1(a) OR (b)

Der Präsident des Europäischen Patentamts;
Im Auftrag

For the President of the European Patent Office

Le Président de l'Office européen des brevets
p.o.

R C van Dijk

BEST AVAILABLE COPY



Anmeldung Nr.:
Application no.: 03000249.7
Demande no:

Anmeldetag:
Date of filing: 08.01.03
Date de dépôt:

Anmelder/Applicant(s)/Demandeur(s):

ARTEMIS Pharmaceuticals GmbH
Neurather Ring 1
51063 Köln
ALLEMAGNE

Bezeichnung der Erfindung/Title of the invention/Titre de l'invention:
(Falls die Bezeichnung der Erfindung nicht angegeben ist, siehe Beschreibung.
If no title is shown please refer to the description.
Si aucun titre n'est indiqué se référer à la description.)

targeted transgenesis using the rosa26 locus

In Anspruch genommene Priorität(en) / Priority(ies) claimed / Priorité(s)
revendiquée(s)

Staat/Tag/Aktenzeichen/State/Date/File no./Pays/Date/Numéro de dépôt:

Internationale Patentklassifikation/International Patent Classification/
Classification internationale des brevets:

G12N15/00

Am Anmeldetag benannte Vertragstaaten/Contracting states designated at date of
filing/Etats contractants désignées lors du dépôt:

AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HU IE IT LU MC NL PT SE SI SK

023403ep JH/BM/ml

Targeted transgenesis using the Rosa26 locus

Introduction

The invention provides a method for targeted transgenesis using the Rosa26 locus. Suitable nucleotide acid sequences and vectors for the targeted transgenesis are provided. The Rosa26 locus proved to be a suitable integration site allowing strong and predictable expression of inserted transgenes carrying exogenous promoters.

Background of the Invention

The generation of transgenic mice by nuclear injection of purified DNA into fertilized eggs is a widely used approach for studying gene or promoter function *in vivo*. However, the level and pattern of expression often varies strongly depending on copy number, configuration, and integration site of the transgene. In addition, founder mice occasionally do not transmit the transgene. Thus, a number of different founders need to be generated and tested in order to identify a useful strain, which is a laborious and time-consuming undertaking (Bradley et al., Nature Genet., 14:121-123 (1996); Jasin et al., Proc. Natl. Acad. Sci. USA, 93:8804-8808 (1996); Dobie et al., Trends Genet., 13:127-130 (1997); Garrick et al., Nature Genet., 18:56-59 (1998), Al-Shawi et al., Mol. Cell. Boil. 10:1192-1198 (1990)).

To overcome these limitations, homologous recombination in embryonic stem cells has been used to produce mice carrying a single copy of the transgene integrated into a predetermined site of the genome (Shaw-White et al., Transgenic Res.; (1):1-13 (1993); Bronson et al., Proc. Natl. Acad. Sci. USA, 93(17):9067-72 (1996); Hatada et al., J. Biol., Chem., 274(2):948-55 (1999); Vivian et al., Biotechniques, 27(1):154-62 (1999); Evans et al., Physiol. Genomics, Mar. 13, 2(2):67-75 (2000); Cvetkovic et al., J. boil. Chem., 275(2):1073-8 (2000); Guillot et al., Physiol. Genomics, Mar. 13, (2):77-83 (2000); Magness et al., Blood, 95(11):3568-77 (2000); Misra et al., BMC Biotechnol., 1(1):12 (2001); Minami et al., Blood, 100(12):4019-25 (2002); Tang et al., Genesis, 32(3):199-202 (2002)). In these studies, the ubiquitous Hprt locus was

2

more or less successfully used for 'targeted transgenesis'. Insertion of a lacZ gene under the control of the polyoma enhancer/HSV thymidine kinase promoter into the third exon of Hprt resulted in variable β -galactosidase expression that was both orientation and cell-type dependent (Shaw-White et al., *Transgenic Res.*; (1):1-13 (1993)). Although transgenes under the control of the human and the chicken β -actin gene promoter resulted in widespread expression when inserted into the Hprt locus, the level of transcripts varied strongly in different tissues (Bronson et al., *Proc. Natl. Acad. Sci. USA*, 93(17:9067-72 (1996)). Unexpectedly, expression of these transgenes, but not of the endogenous Hprt gene appeared to be low or undetectable in kidney and liver (Bronson et al., *Proc. Natl. Acad. Sci. USA*, 93(17:9067-72 (1996)). Hatada et al. demonstrated that the HPRT locus suppresses the activity of both, the haptoglobin gene promoter as well as the herpes simplex thymidine kinase promoter in several tissues of mice (Hatada et al., *J. Biol. Chem.*, 274(2):948-55 (1999)). Likewise, a human eNOS promoter-LacZ reporter gene placed in the Hprt locus was found to be inactive in hepatic vessels that otherwise express the endogenous eNOS gene (Guillot et al., *Physiol. Genomics*, Mar. 13, (2):77-83 (2000). Finally, since the HPRT gene is on the X chromosome, transgene expression at this locus is subjected to random X-inactivation. The expression of the transgene in all cells of the female, therefore, requires the generation of homozygotes.

To avoid the complications referred to above, it would be desirable to define an autosomal locus that allows strong and predictable expression of transgenes inserted through homologous recombination. It is, however, not predictable for a person skilled in the art whether chromosomal loci which fulfill these criteria are available at all. Exogenous transgenes may not harbor all of the sequences necessary and sufficient for proper regulation of transcription and may therefore be influenced by cis-regulatory elements near the site of insertion.

The *rosa26* locus had been identified by random insertion of retroviral sequences and a β -galactosidase-neomycin resistance fusion gene into the genome of mouse embryonic stem cells (Zambrowicz et al., *Proc. Natl. Acad. Sci. USA*, 94, 3789-94 (1997)). The *rosa26* promoter appeared to mediate ubiquitous expression of promoter-less genes

both in embryos and adult mice (Kisseberth et al., Dev. Biol., 214:128-138 (1999); Zambrowicz et al., Proc. Natl. Acad. Sci. USA, 94, 3789-94 (1997)), albeit at different levels in different organs (Voolijs et al., EMBO reports, 21:292-297 (2001)). However, a systematic comparison with other ubiquitous promoters to determine the strength of rosa26 promoter had not been performed. In addition, the activity of exogenous promoters inserted into the rosa26 locus has never been examined.

Summary of the Invention

The present invention is based on the finding that a particular chromosomal locus present within the eukaryotic genome (including that of mammalian ES cells), namely Rosa26, supports the preservation of the inherent activity of heterologous promoters inserted through homologous recombination at that locus. This chromosomal locus is therefore useful in the context of the "targeted transgenesis" approach for the efficient generation of transgenic organisms (such as mice) with a predictable transgene expression pattern.

Such a "targeted transgenesis" method comprises consecutive experimental steps. A gene expression cassette comprising a suitable promoter (e.g. a ubiquitous or tissue specific promoter, either inducible or constitutive) functionally linked to a gene of interest has to be created; subsequently a vector for the targeted insertion of the above mentioned gene expression cassette into the Rosa26 locus has to be generated; the insertion of the above mentioned gene expression cassette into the Rosa26 locus through homologous recombination in embryonic stem cells follows; finally transgenic mice are generated by the injection of such genetically modified ES cells into blastocysts.

More specifically present invention provides

- (1) a method for generating eukaryotic cells having a modified Rosa26 locus, which method comprises the following step (hereinafter shortly referred to as step (a)): introducing a functional DNA sequence into the Rosa26 locus of starting eukaryotic cells by homologous recombination which with a targeting vector comprising said functional DNA sequence flanked by DNA sequences homologous to the Rosa26 locus;

- (2) the method of (1) above, wherein the eukaryotic cells are mammalian embryonic stem (ES) cells, preferably are non-human mammalian ES cells;
- (3) a targeting vector as defined in (1) above;
- (4) eukaryotic cells having a modified Rosa26 locus obtainable by the method of (1) and (2) above;
- (5) a method for preparing a transgenic multi-cell organism having a modified Rosa26 locus which comprises utilizing the method as defined in (1) and (2) above;
- (6) the method of (5) above, wherein the transgenic multi-cell organism is a non-human mammal and said method comprises modifying an ES cell as defined in (2) above;
- (7) a transgenic multi-cell organism and non-human mammal obtainable by the above defined methods (5) and (6), respectively; and
- (8) the use of the eukaryotic cell of (4) above, the transgenic multi-cell organism of (7) above, or the transgenic non-human mammal of (7) above for gene function studies, drug development, as disease model, etc.

The method of the invention offers several advantages over the current technology of pronuclear injection. In particular, the targeting vector allows insertion of a single copy of a gene expression cassette, thus avoiding modulation of transgene expression by the arrangement of multiple copies. By choosing the autosomal Rosa26 locus as insertion site, the expression pattern of the inserted transgene in the non-human animal is predictable; random X-inactivation and/or modulation by chromosomal position effects are avoided. This also eliminates the need to generate and analyse multiple transgenic strains for any given transgene. Finally, the Rosa26 targeting vector for the site-specific integration can be used for multiple gene expression cassettes.

Description of the Figures

Figure 1: Targeted insertion of CreER and CAGGS-Cre-ER into the Rosa26 locus. A cassette comprising a Cre-ER operationally linked to a CAGGS promoter or a cassette comprising a splice acceptor site (SA) linked to a Cre-ER are inserted into the Rosa26 locus via homologous recombination. A perpendicular dash marks the insertion point

within the Rosa26 locus and the rectangular boxes delineate the starting and end points of the Rosa26 transcript.

Figure 2: Southern Blot analysis of the inducible recombination of the Rosa (reporter). (A) Genomic DNA was isolated from liver (Li) spleen (Sp) and small intestine (Si) of transgenic mice carrying the SA-creER/Rosa-rep insert or the CAGGS-creER/Rosa-rep insert. To induce the Cre-ER recombinase the mice were treated with Tamoxifen (treated). As a control, a group of mice with the SA-creER/Rosa-rep insert was left untreated (untreated). Presence of the reporter band (floxed) and deletion (deleted) of it upon an induced recombination event are indicated. (B) Transgenic mice carrying at one Rosa26 locus a loxP flanked DNA polymerase β gene segment ($pol^{\beta flox}$) and at the other a SA-creER/Rosa-rep were treated with Tamoxifen (treated). A control group of mice was left untreated (untreated). Genomic DNA from liver (Li), spleen (Sp), kidney (Ki), heart (He), lung (Lu), thymus (Th), muscle (Mu), small intestine (Si) and brain (Br) was analysed for presence of $pol^{\beta flox}$. In a non-recombination event the $pol^{\beta flox}$ band remained (floxed), In a recombination event deletion occurred (deleted). (C) As (B), but mice carried instead of the SA-creER/Rosa-rep the CAGGS-creER/Rosa-rep insert.

Figure 3: Western Blot analysis of recombinase and α -actin expression. Proteins were extracted from rosa(SA-CreER^{T2}) and rosa (CAGGS-CreER^{T2}) mice and analyzed as described in the "Materials and Method" section. The positions of bands representing CreER^{T2} and actin are indicated. FA: fat tissue, Ty: Thymus; Sp: spleen, Br: Brain, Lu: lung, He: heart.

Detailed Description of the Invention

The term "living organisms" according to the present invention relates to multi-cell organisms which can be vertebrates such as mammals (e.g. non-human animals such as rodents including mice and rats; and humans) or non-mammals (e.g. fish) or can be invertebrates such as insects or worms, or can be plants (higher plants, algi or fungi). Most preferred living organisms are mice and fish.

"Eukaryotic cells" and "starting eukaryotic cells" according to the present invention include cells isolated (derived) from the above defined living organisms and cultured *in*

vitro. These cells can be transformed (immortalized) or untransformed (directly derived from living organisms; primary cell culture). The term "eukaryotic cells" also includes mono-cellular eukaryotic cells such as yeasts, etc.

It is preferred in the method (1) of the present invention that the eukaryotic cells are derived from a multi-cell organism including vertebrates, invertebrates and plants, preferably is a vertebrate cell, more preferably is derived from a mammal, including rodents such as mouse, rat, etc., or a fish such as zebrafish.

In the methods (1) and (2) of the invention it is preferred that the functional DNA sequence comprises a gene encoding a protein/peptide of interest (i.e. is a expressible and translatable DNA sequence), more preferably said functional DNA sequence is a gene expression cassette (a) comprising a gene of interest operatively linked to a promoter, or (b) is a DNA sequence which can be converted into such gene expression cassette (i.e. into an operatively linked "promoter-gene of interest" construct, e.g. by subsequent modification reactions after its integration). The gene of interest within the gene expression cassette can be any gene coding for a certain protein/peptide of interest, including, but not limited to, recombinases, reporter genes, receptors, signaling molecules, transcription factors, pharmaceutically active proteins and peptides, drug target candidates, disease causing gene products, toxins, etc.

The promoter of the gene expression cassette preferably is a ubiquitous or tissue specific promoter, either constitutive or inducible. Preferred ubiquitous promoter is the CAGGS-promoters are CAGGS, hCMV, PGK; preferred tissue specific promoters are FABP (Saam & Gordon, J. Biol. Chem., 274:38071-38082 (1999)), Lck (Orban et al. Proc. Natl. Acad. Sci. USA, 89:6861-5 (1992)), CamKII (Tsien et al., Cell 87: 1317-1326 (1996)), CD19 (Rickert et al., Nucleic Acids Res. 25:1317-1318 (1997)), Keratin (Li et al., Development, 128:675-88 (2001)), Albumin (Postic & Magnuson, Genesis 26:149-150 (2000)), aP2 (Barlow et al., Nucleic Acids Res., 25 (1997)), Insulin (Ray et al., Int. J. Pancreatol. 25:157-63 (1999)), MCK (Brüning et al., Molecular Cell 2:559-569 (1998)), MyHC (Agak et al., J. Clin. Invest., 100:169-179 (1997)), WAP (Utomo et al., Nat. Biotechnol. 17:1091-1096 (1999)), Col2A (Ovchinnikov et al., Genesis

26:145-146 (2000)); preferred inducible promoter systems are. Mx (Kühn et al. Science, 269:1427-1429 (1995)), tet (Urlinger et al., Proc. Natl. Acad. Sci. USA, 97:7963-8 (2000)), Trex (Feng and Erikson, Human Gene Therapy, 10:419-27)

It is moreover preferred that the DNA sequences homologous to the Rosa26 locus are 0.2 to 20 kB, preferably 1 to 10 kB long. The targeting vector, functional DNA sequence or gene expression cassette may further comprise one or more additional functional sequences including but not limited to (selectable) marker genes (such as the neomycin phosphotransferase gene of *E. coli* transposon, etc.), recombinase recognition sites (loxP, FRT, etc.), poly A signals (such as synthetic polyadenylation sites, or the polyadenylation site of human growth hormones, etc.), splice acceptor sequences (such as a splice acceptor of adenovirus, etc.), introns, tags for protein detection, enhancers, selection markers, etc.

In a particularly preferred embodiment of the method (2) the eukaryotic cells are derived from mouse, the DNA sequences homologous to the Rosa26 locus are derived from the 5' and 3' flanking arm of the mouse Rosa26 locus, preferably said homologous DNA sequences having the sequences shown in SEQ ID NO:4 and 5, respectively, and the promoter is a CAGGS-promoter, most preferably the targeting vector has the sequence shown in SEQ ID NO:7.

The methods (1) and (2) may further (besides step (a) defined above) comprise one or more of the steps (b) isolating the eukaryotic cells, preferably the ES cells having the desired functional DNA sequence integrated into the Rosa26 locus; and/or (c) modifying the integrated functional DNA sequence and isolating (ES) cells having the desired modified functional DNA sequence.

The invention also provides a method for preparing a transgenic multi-cell organism having a modified Rosa26 locus which comprises utilizing the method as defined in (1) and (2) above. This includes a method for preparing a non-human mammal comprising modifying starting ES cells according to steps (a) to (c). The ES cells may subsequently be processed according to one or more of the following steps:

(d) the ES cells obtained in steps (b) or (c) are injected into blastocysts; and/or
(e) transgenic non-human animals carrying one or more functional genes of interest at the *Rosa26* locus are generated (viz. by well known breeding procedures).

The transgenic multi-cell organisms and non-human mammals obtainable by the method (5) and (6), respectively; preferably have an operatively functional gene expression cassette (as defined above) integrated into its *Rosa26* locus. Such transgenic multi-cell organisms and non-human mammals are suitable for gene function studies, drug development, as disease model animals, etc.

The invention is further explained by the following examples and the attached figures, which are, however not to be construed so as to limit the invention.

Examples

Materials and Methods

Plasmid construction:

Rosa-targeting vector: A 129 SV/EV-BAC library (Incyte Genomics) was screened with a probe against exon2 of the *Rosa26* locus (amplified from mouse genomic DNA using Rscreen1s (GACAGGACAGTGCTTGTTTAAGG) (SEQ ID NO:1) and Rscreen1as (TGACTACACAATATTGCTCGCAC) (SEQ ID NO:2)). Out of the identified BAC clone a 11 kb *EcoRV* subfragment was inserted into the *HindIII* site of pBS. Two fragments (a 1 kb *SacII/XbaI*- and a 4 kb *XbaI*-fragment) were used as homology arms and inserted into a vector consisting of a FRT-flanked neomycin resistance gene (unpublished) and a splice acceptor site from adenovirus (Friedrich G., Soriano P., *Genes Dev.*, 5:1513-23 (1991)). The CAGGS-promoter or the splice acceptor (SA) and a synthetic intron were inserted between the 5' arm and the neomycin resistance gene. The CreER^{T2} and a polyadenylation site (pA) were cloned 3' of the SA and CAGGS-promoter.

Cell culture: Culture and targeted mutagenesis of ES cells were carried out as previously described (Hogan et al., (Cold Spring Harbor Laboratory Press, Cold Spring Harbor NY.), pp. 253-289.) with ES cell lines derived from both inbred and F1 embryos.

Mice: All mice were kept in the animal facility at Artemis Pharmaceuticals GmbH in microisolator cages (Tecniplast Sealsave). B6D2F1 Mice for the generation of tetraploid blastocysts were obtained from Janvier. The *polb^{fllox}/rosa(CreER^{T2})* and *ect2^{fllox}/rosa(CreER^{T2})* mice were generated by breeding of *rosa(CreER^{T2})* ES mice with BT14 (Gu et al., *Science*, 265, 103-106.), respectively.

Production of ES mice by tetraploid embryo complementation: The production of mice by tetraploid embryo complementation was essentially performed as described (Eggan et al., *Proc Natl Acad Sci USA*, 98, 6209-6214.).

Ligand administration: 100 mg Tamoxifen-free base (Sigma, T5648) was suspended in 100 μ l Ethanol and solved in 1 ml sunflower oil (Sigma). This 10 mg/100 μ l tamoxifen solution was sonicated for 1-2 minutes and then stored at -20°C. For p.o. administration the solution was thawed at 55°C and administrated to 4-8 week old mice by a feeding needle (FST Fine Science Tools GmbH, 18061-20).

Western blot analysis: Western blot analysis was performed using SDS-PAGE (NuPAGE, Invitrogen) and the Breeze Immunodetection System (Invitrogen) according to the manufacturer protocols. Immunodetection was done using sc-543 (HC-20, Santa Cruz Biotechnology, Inc.) against ER, PRB-106C against cre, actin sc-1616 Actin (I-19) against actin and rabbit polyclonal IgG (Santa Cruz Biotechnology, Inc.) antibodies.

Example 1

A CreER^{T2} gene (Feil et al., (1997) *Biochem Biophys Res Commun.*, 237, 752-757) under the control of the CAGGS-promoter (Buchholz et al. (1996), *Nucl. Acids Res.*, 24, 3118-3119) was inserted into the *rosa26* locus by homologous recombination in ES cells (Fig. 1). In addition, the CreER^{T2} gene was introduced along with a splice acceptor sequence (Friedrich and Soziano (1991), *Genes Dev.*, 9, 1513-1523) as a control for

the endogenous activity of the *rosa26* gene promoter (Fig. 1). A *loxP*-flanked hygromycin resistance gene was introduced into the second allele of *rosa26* to provide test substrate for Cre ER^{T2} (Seibler et al., (2003), in press). ES cells modified at both *rosa26* alleles were injected into tetraploid blastocysts and completely ES cell derived mice were generated (Eggan et al., (2001). *PNAS*, 98, 6209-6214). *Rosa*(SA-*CreER* T2 /*reporter*) and *Rosa*(CAGGS-*CreER* T2 /*reporter*) mice were fed with daily 5 mg Tamoxifen for 5 days and recombination of the reporter was analyzed 3 days after the last administration. Southern analysis of genomic DNA from different organs showed up to 50% recombination in the *Rosa*(SA-*CreER* T2 /*reporter*) mice and up to 90% recombination in the *rosa*(CAGGS-*CreER* T2 /*reporter*) mice, respectively (Fig. 2A). As the second substrate, we used the *loxP* flanked DNA *polymerase β* gene segment (*polB* flax) (Gu et al., (1994). *Science*, 265, 103-106). The *polB* flax /*rosa*(SA-*CreER* T2) and *polB* flax /*rosa*(CAGGS-*CreER* T2) mice were fed with 5 mg tamoxifen per day for 5 days and analyzed 3 days later. Southern blot analysis revealed that the *loxP*-flanked *polymerase β* gene segment was excised in more than 90% of cells in all organs except brain in the *rosa*(SA-*CreER* T2 /*reporter*) mice (Fig. 2B). In contrast, the degree of inducible recombination was significantly higher in *rosa*(CAGGS-*CreER* T2 /*reporter*) mice, reaching 100% efficiency in most organs and up to 70% in brain.

To investigate the pattern and level of *CreER* T2 expression in *rosa*(SA-*CreER* T2) and *rosa*(CAGGS- *CreER* T2) mice, we performed Western analysis using antibodies specific for Cre. The 74 kDa band corresponding to the *CreER* T2 fusion protein was detectable in all organs of *rosa*(CAGGS- *CreER* T2) mice, including brain (Fig. 3). In contrast, the *CreER* T2 expression level in *rosa*(SA-*CreER* T2) mice was significantly lower compared to the *rosa*(CAGGS-*CreER* T2) strain and appeared to be undetectable in brain (Fig. 3).

SEQUENCE LISTING

<110> ARTEMIS Pharmaceutical GmbH
 <120> Targeted transgenesis using the Rosa26 locus
 <130> 023403ep JH/BM
 <140>
 <141>
 <160> 7
 <170> PatentIn Ver. 2.1
 <210> 1
 <211> 23
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: primer
 Rscreen1s
 <400> 1
 gacaggacag tgcttggtta agg

23

<210> 2
 <211> 23
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: primer
 Rscreen1s

<400> 2
 tgactacaca atattgctcg cac

23

<210> 3
 <211> 13139
 <212> DNA
 <213> Mus musculus

<220>
 <223> Description: Rosa26 locus

<400> 3
 aagctttctca cgtagcaacc agagctccag agccagcagc tgctgcccgc ttgtatactc 60
 actcctgtga tccaacacag gagcaacctt ttcttttacc cccccccact tcttaacaca 120
 ctttttttttg gggggggggg ggggaacaagt gctocatgct ggaaggattg gaactatgct 180
 tttagaaagg aacaatccta aggtcacttt taaattgagg tcttttgatt gaaaatcaac 240
 aaataccaaa ttccaaatat cgtttttaat taaaccagca atgtggatat aagcattaag 300
 ttttagtttt aaaaagggtc attttccaaa cattcagcaa tcataattta atttacagct 360
 aggaacaaga gccttgggtc atgtcctacc aaagaacata actcaatatt ctacacatga 420
 caatctgaat aaccttaaa cctctaattc cataacaggg cacaaatctt ggacagagaa 480
 ctaatgatcc tcctgagaaa actggaagaa atccagggaa aagaaattcc tgtgtcctcc 540
 aaactcagaa atctctaatt atgtcagtat tctctgcttt agtcctaggt cagattgcac 600
 acatctaaaa taacctctta aagtttttct cctagcgcacc taaaccatta ttaatatcaa 660

attaaccatc	aaaacacttt	cctctcaata	tgctgcacac	aaacctcctc	ctggaaccte	720
ctccatctgg	atcctcccca	atcaaaagta	taggtattta	acatataage	aaggaagtaa	780
tgtaaacatg	accttggtca	caaatatgtc	atctaaaaac	aatttagtca	aggataggag	840
gaaattcgag	aacctgaatc	tttttaagta	ttttgagcac	aggaacaatt	ggcaaaagga	900
atccaggtat	agacaaaacc	cagagcccag	agctctgggc	gaaaaatgag	ttgctggtga	960
agagcctaca	csagttaacat	gagaaagcag	aaaatgcagg	tcattccacg	acccctgacc	1020
caggccagca	gggctgggctg	cagcatcagt	acacaggaga	aagatccctta	ttcctaagaa	1080
tgagaaaggc	aaaggcgccc	gatagaataa	attagcatag	aaggggcttt	cccaggagtt	1140
aaaactttcc	ttctgagcga	ttacctacta	aaaccagggc	ttttgcccac	taccatttac	1200
ctaggatcct	ggcttgacgc	gattcatagg	ggcatatccc	tccccctctt	cttttagagtc	1260
gttcttaaaa	gatcgctctc	cagcccttag	gcagggaaaa	cgacaaaatc	tggctcaatt	1320
ccaggctaga	accctacaaa	ttcaacaggg	atatcgcaag	gatactgggg	catacggcac	1380
agggagtcga	agaatgtgag	gtgggggtgg	cgaaggtaat	gtctttggtg	tgggaaaagc	1440
agcagccatc	tgagatagga	actggaaaaa	cagaggagag	gcgttcagga	agattatgga	1500
ggggaggact	gggccccccac	gagcgaccag	agttgtcaca	aggccgcaag	aacaggggag	1560
gtggggggct	caggggacaga	aaaaaaagta	tgtgtatttt	gagagcaggg	ttgggaggcc	1620
tctcctgaaa	agggatataaa	cgtggagtag	gcaataccca	ggcaaaaagg	ggagaccaga	1680
gtagggggag	gggaagagtc	ctgaccaggg	gaagacatta	aaaaggtagt	gggttcgact	1740
agatgaagga	gagocctttct	ctctgggcaa	gagcggtgca	atgggtgtgta	aaggtagctg	1800
agaagacgaa	aagggcaage	atcttctctg	taccaggctg	gggaggccca	ggcccacgac	1860
cccgaggaga	gggaacgcag	ggagactgag	gtgacccttc	tttccccggg	ggcccggctc	1920
tgtggttcgg	tgtctctctt	ctgttggaac	cttaccttga	cccaggcgct	gcccgggccc	1980
gggcccgggc	tgccggcgac	ggcactcccc	ggaggcagcg	agactcgagt	taggcccac	2040
gcggcgccac	ccoccttctc	ggccgggaat	ggcccgtaac	cgtgaggtgg	gggtgggggg	2100
cagaaaaggc	ggagcgagcc	cgagcgggga	gggggagggc	caggggcgga	ggggggccgg	2160
actactgtgt	tggcggactg	gcgggactag	ggctgcgtga	gtctctgagc	gcaggcgggc	2220
ggcggccgcc	cctccccggg	cggcggcagc	ggcggcagcg	gcggcagctc	actcagccc	2280
ctgcccagac	ggaaaacgca	ctgaccgcac	ggggattccc	agtgcggcg	ccagggggcac	2340
gcgggacacg	ccocctcccc	cccgccatt	ggcctctccg	cccaccgccc	cacacttatt	2400
ggcgggtgcg	ccgccaatca	gcggaggctg	ccggggccgc	ctaaaagaaga	ggctgtgctt	2460
tggggctccg	gctcctcaga	gagcctcggc	taggtagggg	atcgggactc	tggcgggagg	2520
gcggcttggt	gcgttttgcg	ggatgggcgg	ccgcggcagg	ccctccgagc	gtggtggagc	2580
cgttctgtga	gacagccggg	tacgagtcgc	gacgctggaa	ggggcaagcg	ggtggtgggc	2640
aggaatgcgg	tccgcctctg	agcaaccgga	gggggaggga	gaagggagcg	gaaaagtctc	2700
caacggacgc	ggccatggct	cggggggggg	cgggcagcgg	aggagcgctt	ccggccgacg	2760
tctcgtcgct	gattggcttc	ttctcctccc	gccgtgtgtg	aaaacacaaa	tggcgtgttt	2820
tgggttgccg	aaggcgccct	tcagttaacg	gcagccggag	tgcgcagccg	ccggcagcct	2880
cgtcttgccc	actgggtggg	gcgggaggta	ggtgggggtga	ggcagagctg	acgtgcgggc	2940
gcggtcgccc	tctggcgggg	cgggggaggg	gagggagggt	cagcgaaaag	agctcgcgcg	3000
cgagcggccg	cccaccctcc	ccttccctcg	ggggagtcgt	tttaccggcc	gcggccgggg	3060
cctcgtcgct	tgattggctc	tcggggccca	gaaaactggc	ccttgccatt	ggctcgtgtt	3120
cgtgcaagtt	gagtcacatc	gccggccagc	gggggcccgc	aggaggcgct	cccagggttc	3180
ggccctcccc	tgggccccgc	gccgcagagt	ctggccgcgc	gcccctgcgc	aacgtggcag	3240
gaagcgcgcg	ctggggggcg	ggacgggcag	tagggctgag	cggctgcggg	gcgggtgcaa	3300
gcacgtttcc	gacttgagtt	gcctcaagag	gggcgtgctg	agccagacct	ccatcgcgca	3360
ctccggggag	tggaggggaag	gagcgagggc	tcagttgggc	tgttttggag	gcaggaagca	3420
cttgctctcc	caaagtcgct	ctgagttggt	atcagtaagg	gagctgcagt	ggagtaggag	3480
gggagaaggc	cgcacccttc	tccggagggg	ggaggggagt	gttgcaatac	ctttctggga	3540
gttctctgct	gcctcctggc	ttctgaggac	cgcctggggc	ctgggagaat	cccttcccc	3600
tcttccctcg	tgatctgcaa	ctocagtcct	tctagaagat	gggcgggagt	cttctgggca	3660
ggcttaaaag	ctaacctggg	gtgtgggcgt	tgctctgcag	gggaattgaa	caggtgtaaa	3720
attggaggga	caagacttcc	cacagatttt	cggttttgtc	gggaagtttt	ttaatagggg	3780
aaaataagga	aaatgggagg	ataggtagtc	atctgggggt	ttatgcagca	aaactacagg	3840
ttattattgc	ttgtgatccg	cctcggagta	ctttccatcg	aggtagatta	aagacatgct	3900
oacccgagtt	ttatactctc	ctgcttgaga	tccttactac	agtatgaaat	tacagtgtcg	3960
cgagttlagc	tatgttaagca	gaattttaat	catttttaaa	gagcccagta	cttcataatc	4020
atttctcccg	ctccttctgc	agccttatca	aaaggtattt	tagaacactc	attttagccc	4080
catrttcatt	tattatactg	gcttatccaa	cccctagaca	gagcattggc	attttccctt	4140
tcctgatctt	agaagtcctga	tgactcatga	aaccagacag	attagttaca	tacaccacaa	4200
atcgaggctg	tagctggggc	ctcaacactg	cagttctttt	ataactcctt	agtacacttt	4260
ttgttgatcc	tctgccttga	tccttaattt	tcagtgtcta	tcacctctcc	cgtcagtggt	4320

13

gtccacatt	tgggcctatt	ctcagtcag	ggagttttac	aacaatagat	gtattgagaa	4380
tccaacctaa	agcttaactt	tccactccca	tgaatgcctc	tctccttttt	ctccatttat	4440
aaactgagct	attaaccatt	aatgggtcca	gggtggatgct	toctcccat	attacctgat	4500
gtatcttaca	tattgccagg	ctgatatttt	aagacattaa	aaggtatatt	tcattattga	4560
gccacatggt	attgattact	gcttactaaa	attttgtcat	tgtacacatc	tgtaaaagg	4620
ggttcccttt	ggaatgcaaa	gttcagggtg	ttgttgtctt	tcctgacctg	aggtcttgtg	4680
agcttgtatt	ttttctattt	aagcagtgct	ttctcttggg	ctggcttgac	tcattggcatt	4740
ctacacgtta	ttgctggctt	aaatgtgatt	ttgccaagct	tcttcaggac	ctataatttt	4800
gcttgacttg	tagccaaaca	caagtaaaat	gattaagcaa	caaagtattt	tgtgaagctt	4860
ggttcttagg	ttgttgtgtt	gtgtgtgctt	gtgtctctata	ataatactat	ccaggggctg	4920
gagaggtggc	tcggagttca	agagcacaga	ctgctcttcc	agaagtcttg	agttcaattc	4980
ccagcaacca	catgggtggc	cacaaccatc	tgtaatggga	tctgatgccc	tcttctgggtg	5040
tgtctgaaga	ccacaagtgt	attcacatta	aataaataaa	tctccttctt	tcttcttctt	5100
ttttttttta	aagagaatac	tgtctccagt	agaatttact	gaagtaatga	aatactctgt	5160
gtttgttcca	atatggtagc	caataatcaa	attactcttt	aagcactgga	aatgttacca	5220
aggaactaat	ttttatttga	agtgttaactg	tggacagagg	agccataact	gcagacttgt	5280
gggatacaga	agaccaatgc	agactttaat	gtcttctctc	ttacactaag	caataaagg	5340
ataaaaaattg	aacttctagt	atcctatttg	tttaaactgc	tagctttact	taacttttgt	5400
gcttcatcta	tacaaagctg	aaagctaagt	ctgcagccat	tactaaacat	gaaagcaagt	5460
aatgataatt	ttggatttca	aaaatgtagg	gccagagttt	agccagccag	tgggtgggtg	5520
tgccctttatg	ccttttaatcc	cagcactctg	gaggcagaga	caggcagatc	tctgagtttg	5580
agcccagcct	gggtctacaca	tcaagttcta	tctaggatag	ccaggaatac	acacagaaac	5640
cctgttgggg	agggggggctc	tgagatttca	taaaattata	attgaagcat	tccctaatag	5700
gccactatgg	atgtggctaa	atccgtotac	ctttctgatg	agatttgggt	attatrtttt	5760
ctgtctctgc	tggttgggtg	gtctrttgac	actgtgggct	ttctttaaag	cctccttctt	5820
gccatgtggg	ctcttgtttg	ctactaactt	cccatggctt	aaatggcatg	gctttttgct	5880
ttctaagggc	agctgctgag	atthgcagcc	tgatttccag	gggtggggtg	ggaaatcttt	5940
caaacactaa	aattgtctctt	taattttttt	tttaaaaaat	gggttatata	ataaacctca	6000
taaaatagtt	atgaggagtg	aggtgggacta	atattaaatg	agtcctctcc	ctataaaaaga	6060
gctattaagg	ctttttgtct	tataotaaac	ttttttttta	aatgtgggat	ctttagaacc	6120
aagggtctta	gagtttttagt	atacagaaac	tggtgcatcg	cttaatcaga	ttttctagtt	6180
tcaaatccag	agaatccaaa	ttcttcacag	ccaaagtcaa	attaagaatt	tctgactttt	6240
aatgttaatt	tgettactgt	gaatataaaa	atgatagett	ttcctgaggg	aggggtctcac	6300
tatgtatctc	tgcttgatct	gcaacaagat	atgtagacta	aagtctctgc	tgcttttgtc	6360
tcctgaatac	taagggtaaa	atgtagtaat	acttttggaa	crtgcaggct	agattctttt	6420
ataggggaca	cactaaggga	gcttgggtga	tagttgggtaa	aatgtgtttc	aagtgtgaa	6480
aacttgaatt	actatcccg	caacotactt	tttaaaaaaa	aaagccaggc	ctgttagagc	6540
atgotttaagg	gatcctcagg	acttgctgag	cacacaagag	tagttacttg	gcaggctctc	6600
gggtgagagca	tatttcaaaa	aacaaggcag	acaaccaaga	aactacagtt	aagggtacct	6660
gtcttttaaac	catctgcata	tacacaggga	tattaaataa	ttccaaataa	tatttcatcc	6720
aagtrttccc	ccatcaaaat	gggacatgga	ttctctcggt	gaataggcag	agttggaaaac	6780
taaacaaaatg	ttgggttttgt	gatttgtgaa	attgtrttca	agtgatagtt	aaagcccatg	6840
agatacagaa	caaagctgct	atthtcgaggt	ctcttggttt	atactcagaa	gcacttcttt	6900
gggtttccct	gcactatcct	gatcatgtgc	taggcotacc	ttaggctgat	tggtgttcaa	6960
ataaaacttaa	gtttccctgct	aggtgatgtc	atatgatttc	atatatcaag	gcaaaacatg	7020
ttatatatgt	taaacatttg	tacttaatgt	gaaagttagg	tctttgtggg	tttgattttt	7080
aattttcaaa	acctgagcta	aataagteat	ttttacatgt	cttacatttg	gtggaaattgt	7140
ataattgtgg	tttgagggca	agactctctg	acctagtaac	cctacctata	gagcattttg	7200
ctgggtcaca	agtctaggag	tcaagcattt	caccttgaag	ttgagacgtt	ttgttagtgt	7260
atactagttt	atatgttggg	ggacatgttt	atccagaaga	tattcaggac	tatttttgac	7320
tgggttaagg	aattgattct	gattagcact	gttagtgagc	attgagtggc	ctttaggctt	7380
gaattggagt	cacttgtata	tctaaaataa	tgctggcctt	ttttaaaaag	cccttgttct	7440
ttatcacctt	gtttttctaca	taatttttgt	tcaaagaaat	acttgtcttg	atctcctttt	7500
gacaacaata	gcatgttttc	aagccatatt	ttttttctct	tttttttttt	tttttggttt	7560
ttcgagacag	ggttttctctg	tatagccctg	gctgtctctg	aactcacttt	gtagaccagg	7620
ctggcctcga	actcagaaat	ccgcctgcct	ctgcctctg	agtgcgggga	ctaaaggcgt	7680
gcaccaccac	gcctggctaa	gttggatatt	ttgttatata	actataacca	atactaactc	7740
cactgggtgg	atttttaatt	cagtcagtag	tcttaagtgg	tctttatttg	cccttcatta	7800
aaatctactg	ttcactctaa	cagaggctgt	tggtactagt	ggcacttaag	caacttctta	7860
cggatatact	agcagattaa	gggtcaggga	tagaaaactag	tctagcgttt	tgtataccta	7920
ccagctttat	actacettgt	tctgatagaa	atatttcagg	acacttagag	tgtactataa	7980

14

ggttgatggg	aagcttataa	ggaacttgaa	agtggagtaa	ctactccatt	tctctgaggg	8040
gagaattaaa	atttttgacc	aagtgtgtgt	gagccactga	gaatgggtctc	agaacataac	8100
ttcttaagga	accttcccag	attgccctca	acactgcacc	acatttggtc	ctgcttgaac	8160
attgccatgg	ctcttaaggt	cttaattaag	aatattaatt	gtgtaattat	tgtttttcct	8220
ccttttagatc	attccttgag	gacaggacag	tgcttggtta	aggetatatt	tctgctgtct	8280
gagcagcaac	aggctcttga	gatcaacatg	argttcataa	tcccaagatg	ttgccattta	8340
tggtctcaga	agcaagcaga	ggcatgatgg	tcagtgacag	taatgtcact	gtgttaaattg	8400
ttgctatgca	gtttggattt	ttctaattgta	gtgtaggtag	aacatatgtg	ttctgtatga	8460
attaaaotct	taagttacac	cttgrataat	ccatgcaatg	tgttatgcaa	ttaccatttt	8520
aagtattgta	gctttctttg	tatgtgagga	tcaagggtgt	tgtcataaaa	tgttttgaac	8580
atttccccaa	agttccaaat	tataaaacca	caacgttaga	acttattttat	gaacaatggt	8640
tgragtttca	tgctttttaa	atgcttaatt	attcaattaa	caccgtttgt	gttataatat	8700
atataaaaact	gacatgtaga	agtgtttgtc	cagaacattt	cttaaatgta	tactgtcttt	8760
agagagttta	atatagcatg	tcttttgcaa	catactaact	tttgtgttgg	tgcyagcaat	8820
attgtgtagt	cattttgaaa	ggagtcattt	caatgagtg	cagattgttt	tgaatgttat	8880
tgaacatttt	aaatgcagac	ttgttcgtgt	tttagaaagc	aaaaactgtca	gaagctttga	8940
actagaaaat	aaaaagctga	agtatttcag	aagggaaata	agctacttgc	tgtattagtt	9000
gaaggaaaag	gtaatagctt	agaaaattta	aaaccatata	gttgtcattg	ctgaatatct	9060
ggcagatgaa	aagaaatact	cagtgggtct	tttgagcaat	ataacagctt	gttatattaa	9120
aaattttccc	cacagatata	aactctaact	tataactcat	aaatgttaca	aatggatgaa	9180
gcttaciaaat	gtggcttgac	ttgtcactgt	gcttgcttta	gttatgrgaa	agtttggaac	9240
taaacctatg	tcctaataatg	tcaaaactgtg	gaatgacttt	ttaatctatt	ggtttgtcta	9300
gaacagttat	gttgccattt	gcctaataatg	tgaagaaaaa	agtggggagt	gccttggcac	9360
tggtcatttg	tggtgtgaac	caaagagggg	ggcatgcact	tacacttcaa	acatcctttt	9420
gaaagactga	caagtttggg	tcttcacagt	tggaattggg	catccctttt	gtcaggagg	9480
gagggagggg	gggaggctgg	cttgttatgc	tgacaagtgt	gattaaattc	aaactrtgag	9540
gtaagttgga	ggaacttgta	cattgttagg	agtgtgacaa	tttggactct	taatgatttg	9600
gtcatacaaa	atgaacctag	accaacttct	ggaagatgta	tataataact	ccatgttaca	9660
ttgatttcac	ctgactaata	cttatccctt	atcaattaaa	tacagaagat	gccagccatc	9720
tgggcctttt	aaccagaaaa	tttagtttca	aactoctagg	ttagtgttct	cactgagcta	9780
catcctgate	tagtcttgaa	aataggacca	ccatcacccc	caaaaaaatc	tcaaataga	9840
tttatgctag	tgtttcaaaa	tttttaggaat	aggtaagatt	agaaagtttt	aaattttgag	9900
aaatggcttc	tctagaaaga	tgtacatagt	gaacactgaa	tggtctcctaa	agagcctaga	9960
aaactggtag	tgagcacaca	ggactgagag	gtctttcttg	aaaagcatgt	attgctttac	10020
gtgggtcaca	gaaggcaggc	aggaagaact	tgggctgaaa	otgggtgtctt	aagtggctaa	10080
catcttcaca	actgtgagc	aagaacttta	tcctgatgca	aaaaccatcc	aaacaaacta	10140
agtgaagggt	ggcaatggat	cccaggctgc	tctagaggag	gacttgactt	ctcatcccat	10200
caccacacac	agatagctca	tagactgcca	attaacacca	gcttctagcc	tccacaggca	10260
cctgcactgg	tacacataat	ttcacacaaa	cacagtaaga	agccttccac	ctggcatggg	10320
attgcttate	tttagttccc	aacacttggg	aggcagaggc	cagccagggc	tatgtgacaa	10380
aaaccttgct	tagaggagaa	acttcatagc	ttatcttcta	ttcaogtaac	caggtagtga	10440
aaattttacca	gccagagatg	aagctaacag	tgtccactat	atttgtagt	ttttaagtca	10500
attttttaaa	tatacttaat	agaattaaag	ctatggtgaa	ccaagtacaa	acctggtgta	10560
ttaaacttgag	aaacttagcat	aaaaagtagt	tcatttgttc	agtaaatatt	aaatgcttac	10620
tggcaagat	tatgtcagga	acttggtaaa	tgggtgatgaa	acaatcatag	ttgtacatct	10680
tggttctgtg	atcaccttgg	tttgaggtaa	aagtgggtcc	tttgatcaag	gatggaattt	10740
taagtttata	ttcaatcaat	aatgtattat	tttgbattg	caaaattgce	tacttagggg	10800
ataaaaocctt	taaaaatttc	ataataccag	ttcattctcc	agttactaat	tccaaaaagc	10860
cactgactat	ggtgccaatg	tggattctgt	tctcaaagga	aggattgtct	gtgcccttta	10920
ttctaataga	aacatcacac	tgaaaaatcta	agctgaaaga	agccagactt	tcctaaataa	10980
ataactttcc	ataaagctca	aacaaggatt	acttttagga	ggcactgtta	aggaactgat	11040
aagtaatgag	gttacttata	taatgatagt	cccacaagac	tatctgagga	aaaatcagta	11100
caactcgaaa	acagaacaac	cagctaggga	ggaataacag	ggctcccaag	ccaggagggtc	11160
tatccaacac	ccttttctgt	tgaggggccc	agacctacat	attgtataca	aacagggagg	11220
tgggtgattt	taactctcct	gagggtacott	ggtaaatctt	tgtcctgagt	aagcagtaca	11280
gtgtacagtt	tacattttca	tttaaagata	cattagctcc	ctctaccccc	taagactgac	11340
aggcactttg	gggggtgggga	gggctttgga	aaataacgct	tccatacact	aaaagagaaa	11400
ttcttttaat	taggtctgtt	ggttccatc	attcactggt	gtttctacta	cttagtaata	11460
ttataatagt	cacacaagca	tctttgtctt	gtttagggtg	tatatttatt	ttaaaggcaga	11520
tgataaaaact	gtagatctta	agggatgctt	ctgcttctga	gatgatacaa	agaatttaga	11580
ccataaaaaca	gtaggttgca	caagcaatag	aatatggcct	aaagtgttct	gacacttaga	11640

15

agccaagcag	tgtaggcttc	ttaagaaata	ccattacaat	caccttgcta	gaaatcaagc	11700
attctggagt	ggccaagcag	tgtaacctgt	actgtaagtt	acttttctgc	tatttttctc	11760
ccaaagcaag	ttctttatgc	tgatatttcc	agtgttagga	actacaaata	tttaataagtt	11820
gtottcactc	ttttctttac	caaggagggt	ctcttctctc	atcttgatct	gaaggatgaa	11880
caaaggcttg	agcagtgcgc	tttagaagat	aaactgcagc	atgaaggccc	ccgatgttca	11940
cccagactac	atggaccttt	cgccacacat	gtcccattcc	agataaggcc	tggcacacac	12000
aaaaaacata	agtcattagg	ctaccagtct	gattctaaaa	caacctaaaa	tcttcccact	12060
taaattgctat	gggtgggtggg	ttggaaagtt	gactcagaaa	atcactcgct	gtttttagag	12120
aggatctggg	ttcagtttct	gatacattgt	ggcttacaac	tataactcca	gttctagggg	12180
gtccatccaa	catcctcttc	tggtgagggc	accaaataaa	tgtattgtgt	acaaacaggg	12240
aggtgagtga	tttaactctc	gtgtatagta	ccttggtaaa	acatttcttg	tcttgagtaa	12300
gcagtacagc	tctgcctgtc	cctgggtctac	agacacggct	catttcccga	aggcaagctg	12360
gatagagatt	ccaatttctc	ttcttggatc	ccatcctata	aaagaaggtc	aagtttaatc	12420
tattgcaaaa	ggtaaatagg	tagtttctta	catgagacaa	gaacaaatct	taggtgtgaa	12480
gcagtcattc	ttacaggcc	agagcctcta	ttctatgcca	atgaaggaaa	ctgttagtcc	12540
agtgttatag	agttagtcca	gtgtatagtt	ttctatcaga	acactttttt	tttaaacac	12600
tgcaacttag	cttattgaag	acaaaccacg	agtagaaatc	tgtccaagaa	gcaagtgtct	12660
ctcagcctac	aatgtggaat	aggaccatgt	aatggtacag	tgagtgaat	gaattatggc	12720
atgtttttct	gactgagaag	acagtacaat	aaaaggtaaa	ctcatgggat	ctatttaaaa	12780
agaatccaat	ttctaccttt	ttccaaatgg	catatctggt	acaataatat	ccacagaagc	12840
agttctcagt	gggagggtgc	agatatccca	ctgaacagca	tcaatgggca	aaccccaggt	12900
tgtttttctg	tgagagacaaa	ggtaagatat	ttcaatatat	tttcccagc	taatgagatg	12960
gctcagcaaa	taatgggtact	ggccatttaag	tctcatgacc	tgagcttgat	cctcagggac	13020
catgtggtac	aaggagagac	ctaaatcctt	cagttggact	tcaatcttct	accctcatgt	13080
ccacacacaa	ataaatacaa	taaaaaacat	tctgcagtcg	aatttctaaa	agggcgaat	13139

<210> 4
 <211> 1073
 <212> DNA
 <213> Mus musculus

<220>
 <223> Description: 5' arm for Rosa26

<400> 4	
caggccctcc	gagcgtgggtg
ggaaggggca	agcgggtgggt
gggagaaggg	agcggaaaaag
gcggaggagc	gcttccgggc
tgtgaaaaca	caaattggcgt
ggagtgcgca	gcgcgcggca
gtgagggcag	ctggacgtgc
gggtcagcga	aagttagctcg
tctgttttacc	cgcgcgcggc
tggtcccttgc	cattggctcg
ggcggaggagg	cgctcccagg
gcgcgccccct	gcgcaacgtg
tgagcggcctg	cggggcggggt
gctgagccag	acctccatcg
gggtctgtttt	ggaggcagga
aaggggagctg	cagtggagta
gagtgttgca	ataaccttct
gggcctggga	gaatcccttc
gagccgtctct	gagccgtctct
gtgagacagc	gtgagacagc
cggttccgcc	cggttccgcc
acgcggccat	acgcggccat
cgctgattgg	cgctgattgg
gcgtaaggcg	gcgtaaggcg
cctgtcagtt	cctgtcagtt
tggggcgggga	tggggcgggga
ggtaggtggg	ggtaggtggg
agggggagggg	agggggagggg
ctcccccttc	ctcccccttc
gctctcgggg	gctctcgggg
atccgcgggc	atccgcgggc
cgcgcccgca	cgcgcccgca
gcggggacgg	gcggggacgg
agttgcctca	agttgcctca
gaaggagcga	gaaggagcga
cgctctgagt	cgctctgagt
cttctccgga	cttctccgga
ggggggagggg	ggggggagggg
ggaccgccct	ggaccgccct
tct	tct

<210> 5
 <211> 4333
 <212> DNA
 <213> Mus musculus

<220>

<223> Description: 3' arm for Rosa26

<400> 5

tagaagatgg	gcgggagtc	tctgggcagg	cttaaaggct	aacctggtgt	gtgggcgttg	60
tcctgcaggg	gaattgaaca	ggtgtaaaat	tggaggggaca	agacttccca	cagattttcg	120
gttttgtcgg	gaagtttttt	aataggggca	aataaggaaa	atgggaggat	aggtagtcat	180
ctgggggtttt	atgcagcaaa	actacagggt	attattgctt	gtgatccgcc	tcggagtatt	240
ttccatcgag	gtagattaaa	gacatgctca	cccaggtttt	atactctcct	gcttgagatc	300
cttactacag	tatgaaatta	cagtgtcgcg	agttagacta	tgtaagcaga	attttaatca	360
tttttaaaaga	gcccagttact	tcatatccat	ttctcccgc	ccttctgcag	ccttatcaaa	420
aggattttta	gaacactcat	tttagcccca	ttttcattta	ttatactggc	ttatccaacc	480
cctagacaga	gcattggcat	tttccctttc	ctgatccttag	aagtctgatg	actcatgaaa	540
ccagacagat	tagttacata	caccacaaat	cgaggctgta	gctggggcct	caacactgca	600
gttcttttat	aactccttag	cacacttttt	ggtgatcctt	tgctctgatc	cttaattttc	660
agtgtctatc	acctctcccg	tcagtgggtg	tccacatttg	ggcctattct	cagtcacagg	720
agtttttcaaa	caatagatgt	attgagaatc	caacctaaag	cttaactttc	cactcccatg	780
aatgcctctc	tcctttttct	ccatttataa	actgagctat	taaccattaa	tggttccagg	840
tggatgtctc	ctcccatat	tacctgatgt	atcttacata	ttgccaggct	gatattttta	900
gacattaaaa	ggtatatatt	attattgagc	cacatgggtat	tgattactgc	ttactaaaaa	960
tttgtcattg	tacacatctg	taaaagggtg	ttccttttgg	aatgcaaggt	tcagggtgtt	1020
ggtgtctttt	ctgacctaa	gtcttgtgag	cttgtatttt	ttctatttaa	gcagtgtctt	1080
ctcttggact	ggcttgactc	atggcattct	ccacgtttat	gctgggtctaa	atgtgatttt	1140
gccaagcttc	ttcaggacct	ataattttgc	ttgacttgta	gccaacacac	agtaaaatga	1200
ttaagcaaca	aatgtatctg	tgaagcttgg	tttttaggct	gttgtgttgt	gttgtgttgt	1260
gctctataat	aatactatcc	aggggctgga	gagggtggctc	ggagtccaag	agcacagact	1320
gctcttccag	aagtccctgag	ttcaattccc	agcaaccaca	tggtggctca	caaccatctg	1380
taattgggatc	tgatgccctc	ttctgggtgt	tctgaagacc	acaagtgtat	tcacattaaa	1440
taaataaaatc	ctccttcttc	ttcttttttt	ttttttttaa	gagaatactg	tctccagtag	1500
aatttactga	agtaatgaaa	tacttttgtg	ttgttccaat	atggtagcca	ataatcaaat	1560
tactcttttaa	gcactggaaa	tgttaccaag	gaactaattt	ttatttgaag	tgtaactgtg	1620
gacagaggag	ccataactgc	agacttgtgg	gatacagaag	accaatgcag	actttaactg	1680
cttttctctt	acactaagca	ataaagaaat	aaaaattgaa	cttctagtat	cctatttgtt	1740
taaaactgcta	gctttactta	acttttgtgc	ttcatctata	caaagctgaa	agctaagtct	1800
gcagccatta	ctaaacatga	aagcaagtaa	tgataatttt	ggatttcaaa	aatgttagggc	1860
cagagtttag	ccagccagtg	gtgggtgctg	ccttttatgcc	tttaatccca	gcactctgga	1920
ggcagagaca	ggcagatctc	tgagtttgag	cccagcctgg	tctacacatc	aagttctate	1980
taggatagcc	aggaatacac	acagaaaccc	tgttggggag	gggggctctg	agatttcata	2040
aaattataat	tgaagcattc	cctaattgagc	cactatggat	gtggctaaat	cogtctacct	2100
ttctgatgag	atttgggtat	tattttttct	gtctctgctg	ttgggtgggt	cttttgacac	2160
tgtgggcttt	ctttaaagcc	tccttctctg	catgtgggtc	cttggttgct	actaaacttc	2220
catggcttaa	atggcatggc	tttttgcctt	ctaagggcag	ctgctgagat	ttgcagcctg	2280
atttccaggg	tggggttggg	aaatctttca	aacactaaaa	ttgtccttta	attttttttt	2340
taaaaaatgg	gttatataat	aaacctcata	aaatagtctt	gaggagttag	gtggactaat	2400
attaaatgag	tccttccctt	ataaaagagc	tattaaggct	ttttgtctta	tacttaactt	2460
ttttttttaa	tgtggtatct	tcagaaccaa	gggtccttaga	gttttagtat	acagaaactg	2520
ttgcategct	taatcagatt	ttctagtttc	aaatccagag	aatccaaatt	cttcacagcc	2580
aaagtcaaat	taagaatttc	tgacttttaa	tgtaattttg	cttactgtga	atataaaaaa	2640
gatagctttt	cctgaggcag	ggtctcacta	tgatctctct	cctgatctgc	aacaagatat	2700
gtagactaaa	gttctgcctg	cttttgtctc	ctgaatacta	aggttaaaaa	gtagtaatac	2760
ttttgggaact	tgagggtcag	attcttttat	ctgggacaca	ctaaggggagc	ttgggtgata	2820
gttgggtaaaa	tgtgtttcaa	gtgatgaaaa	cgtgaactat	tatcacggca	acctactttt	2880
taaaaaaaaa	agccaggcct	gttagagcat	gcttaaggga	tccttaggac	ttgctgagca	2940
cacaagagta	gttacttggc	aggctcctgg	tgagagcata	tttcaaaaaa	caaggcgagc	3000
aaccaagaaa	ctacagttaa	ggttacctgt	ctttaaacca	tctgcatata	cacaggggata	3060
ctaaaaatatt	ccaaaataata	ttcoattcaa	gttttccccc	atcaaattgg	gacattggatt	3120
totccgggtga	ataggcgagag	ttggaaacta	aacaaatgtt	ggttttgtga	tttgtgaaat	3180
tggttttcaag	tgatagttaa	agcccatgag	atacagaaca	aagctgctat	ttcgaggtct	3240
cttgggtttat	actcagaagc	actctcttgg	gtttccctgc	actatcctga	tcattgtgcta	3300
ggcctacett	aggetgatgg	ttgttcaaat	aaacttaagt	ttcctgtcag	gtgatgtcat	3360
atgattttcat	atatcaaggc	aaaacatgct	atatatgtta	aacatttgta	cttaattgtga	3420

17

```

aagttagggtc tttgtgggtt tgattttttaa ttttcaaaac ctgagctaaa taagtcattt 3480
ttacatgtct tacatttggg ggaattgtat aattgtgggt tgcaggcaag actctctgac 3540
ctagtaaccc tacctataga gcactttgct gggtcacaag tctaggagtc aagcatttca 3600
ccttgaagtt gagacgtttt gttagtgtat actagtttat atgttggagg acatgtttat 3660
ccagaagata ttcaggacta tttctgactg ggctaaggaa ttgattctga ttagcactgt 3720
tagtgagcat tgagtggcct ttaggcttga attggagtca cttgtatate tcaataatg 3780
ctggcctttt ttaaaaagcc cttgttcttt atcaccctgt tttctacata atttttgttc 3840
aaagaaatac ttgtttggat ctcccttttg caacaatagc atgttttcaa gccatatttc 3900
tttccctttt tttttttttt tttgggtttt cgagacaggg tttctctgta tagccctggc 3960
tgtcctggaa ctccacttgt agaccaggct ggctcgaac tcagaaatcc gctgctct 4020
gcctcctgag tgcggggatt aaaggcgtgc accaccagc ctgggtggat tggatatttt 4080
gttatataac tataaccaat actaactcca cttcattaaa atctactgtt cttcaattca gtcagtagtc 4140
ttaagtggtc tttattggcc ctccattaaa gatatactag cactctaaca gaggctgttg 4200
gtactagtgg caactaagca acttcctacg gatatactag cagattaagg gtcagggata 4260
gaaactagtc tagcgttttg tatacctacc agctttatac taccttgttc tgatagaaat 4320
atttcaggac atc

```

4333

<210> 6

<211> 6039

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Insert
CAGGAS-creER

<400> 6

```

attgattatt gactagttat taatagtaat caattacggg gtcattagtt catagcccat 60
atatggagtt ccgcgttaca taacttacgg taaatggccc gcctggctga ccgcccacg 120
acccccgccc attgacgtca ataatgacgt atgttcccat agtaacgcca atagggactt 180
tccattgacg tcaatgggtg gactatttac ggtaaaactgc ccacttggca gtacatcaag 240
tgtatcatat gccaaagtacg cccctattg acgtcaatga cggtaaatgg cccgcctggc 300
attatgcccc gtacatgacc ttatgggact ttcctacttg gcagtacatc tacgtattag 360
tcactgctat taccatgggt cgaggtgagc cccacgttct gcttcactct ccccatctcc 420
ccccctccc caccocccat tttgtattta tttatttttt aattattttg tgcagcgatg 480
ggggcggggg gggggggggc gcgcgccagg cggggcgggg cggggcgagg ggcggggcgg 540
ggcgaggcgg agaggtgcgg cggcagccaa tcagagcggc gcgctccgaa agtttccctt 600
tatggcgagg cggcgggcgg ggcggcccta taaaaagcga agcgcgcggc gggcgggagt 660
cgctgcgttg ccttcgcccc gtgccccgct ccgcgcgcgc tcgcgcgcgc cgccccggct 720
ctgactgacc gcgttactcc cacaggtgag ogggcgggac ggcccttctc ctccgggctg 780
taattagcgc ttggtttaat gacggctcgt ttctttcttg tggctgcgtg aaagccttaa 840
agggctccgg gagggccctt tgtgccccgg ggagcggctc ggggggtgag gctgcggggc 900
tgtgcgtggg gagcgccgag tgccggccgc gctgcccggc ggctgtgagc gctgcggggc 960
cggcgcgggg ctttgtgcgc tccgcgtgtg gcgaggggga gcgcgccggg gggcggtgac 1020
ggtaggcagg ggggtgtggc gcggcggtcg ggctgttaac cccccctgca cccccctccc 1140
cgagttgctg agcacggccc ggcttcgggt gcggggctcc gtgcggggcg tggcgcgggg 1200
ctcgccgtgc cgggcggggg gtggcggcag gtgggggtgc cgggcggggc ggggcccgcct 1260
cggggccggg agggctcggg ggagggggcg ggccgcccgg gagcgccggc ggctgtcgag 1320
gcgcggcgag ccgcagccat tgccctttat ggtaatcgtg cgagagggcg cagggaactc 1380
ctttgtccca aatctggcgg agccgaaatc tgggaggcgc cgcgcacccc cctctagcgg 1440
gcgcggggcg agcgggtgcgg ccgcggcagg aaggaaatgg gcggggaggg ccttcgtgcg 1500
tcgcgcgcgc gccgtccctc tctccatctc cagcctcggg gctgcccag ggggacggct 1560
gccttcgggg gggacggggc agggcggggt tcggcttctg gcgtgtgacc ggcggctcta 1620
gaagcgttgg ggtgagtact ccctctcaaa agcgggcgat acttotgcgc taagattgtc 1680
agtctccaaa aacgaggagg atttgatatt caccctggcc gcggtgagc ctttgagggt 1740
ggcgcgtcc atctgggtcag aaaagacaat ctttttgggt tcaagcttga ggtgtggcag 1800
gcttgagtc tggccataca cttgagtgac attgacatcc actttgcctt tctotccaca 1860
ggtgtccact cccagggcgg cctccggagc gatcgccggg ccgcctaggc tagccggcgg 1920
cgctcgaccat gtccaattta ctgaccgtac accaaaactt gcctgcatta ccgggtcgatg 1980

```

caacgagtgatga	tgagggttcgc	aagaacctga	tggacatggt	cagggatcgc	cagggcgtttt	2040
ctgagcatatc	ctgggaaatg	cttctgtccg	tttgccgggtc	gtgggcggca	tggtgcaagt	2100
tgaataaccg	gaaatgggtt	cccgcagaac	ctgaagatgt	tcgcgattat	cttctatatc	2160
ttcagggcg	cggtctggca	gtaaaaacta	tcagcaaca	tttggggccag	ctaaacatgc	2220
ttcatcgctcg	gtccgggctg	ccacgaccaa	gtgacagcaa	tgctgtttca	ctgggttatgc	2280
ggcggtatccg	aaaagaaaac	gttgatgccg	gtgaacgtgc	aaaacaggct	ctagcgttcg	2340
aacgcactga	tttcgaccag	gttcgttcac	tcattggaaaa	tagcgatcgc	tgccaggata	2400
tacgtaatct	ggcattttctg	gggattgctt	ataacaccct	gttacgtata	gccgaaattg	2460
ccaggatcag	ggtraaagat	atctcacgta	ctgacgggtg	gagaatgtta	atccatattg	2520
gcagaacgaa	aacgctgggt	agcaccgcag	gtgtagagaa	ggcacttagc	ctgggggttaa	2580
ctaaactgggt	cgagcgatgg	atttccgtct	ctgggtgtagc	tgatgatccg	aataactacc	2640
tggttttgccg	ggtcagaaaa	aatgggtgtg	ccgcgccatc	tgccaccagc	cagctaccaa	2700
ctcgcgccct	ggaagggatt	tttgaagcaa	ctcatcgatt	gatttacggc	gctaaggatg	2760
actctgggtca	gagatacctg	gcctgggtctg	gacacagtg	ccgtgtcggg	gccgcgcgag	2820
atatggcccg	cgctggagtt	tcaataccgg	agatcatgca	agctgggtgg	tggaaccaatg	2880
taaatatgt	catgaactat	atccgtaaac	tggatagtga	aacaggggca	atgggtgcgc	2940
tgctggaaga	tggcgattct	gctggagaca	tgagagctgc	caaccttgg	ccaagcccg	3000
tcattgatcaa	acgctctaag	aagaacagcc	tggccttgtc	cctgacggcc	gaccagatgg	3060
tcagtgcctt	gttggatgct	gagcccccac	tactctattc	cgagtatgat	cctaccagac	3120
ccttcagtga	agcttcgatg	atgggcttac	tgaccaacct	ggcagacagg	gagctgggtc	3180
acattgatcaa	ctggggcgag	aggggtgccag	gctttgtgga	tttgaccctc	catgatcagg	3240
tccaccttct	agaatgtgcc	tggctagaga	tctgatgat	tggtctcgtc	tggcgctcca	3300
tggtgaccc	agtgaagcta	ctgtttgtct	ctaaactgct	cttgagacagg	aaccagggaa	3360
aatgtgtaga	gggcatgggt	gagatcttcg	acatcgtgct	ggctacatca	tctcggttcc	3420
gcattgatgaa	tctgcaggga	gaggagtgtg	tgtgctcaa	atctattatt	ttgcttaatt	3480
ctggagtgt	cacattttctg	tccagcacc	tgaagtctct	ggaagagaag	gaccatatcc	3540
accgagtcct	ggacaagatc	acagacactt	tgatccacct	gatggccaag	gcaggcctga	3600
ccctgcagca	gcagcaccag	cggctggccc	agctcctcct	catcctctcc	cacatcaggc	3660
acattgagtaa	caaaggcatg	gagcatctgt	acagcatgaa	gtgcaagaac	gtgggtgccc	3720
tctatgacct	gctgctggag	gcggccgacg	cccaccgcct	acatgcgccc	actagccgtg	3780
gagggggcatc	cgtggaggag	acggaccaaa	gccacttggc	cactgcgggc	tctacttcat	3840
cgcatctcct	gcaaaagtat	tacatcacgg	gggaggcgaga	gggtttccct	gccacagtct	3900
gagcggccga	ccgggttcgag	atccaggcgc	ggatcaataa	aagatcatta	ttttcaatag	3960
atctgtgtgt	tggttttttg	tgtgccttgg	gggagggggga	ggccagaatg	aggcgcgggc	4020
aaggggggag	gggaggccag	aatgaccttg	ggggaggggg	aggccagaat	gaccttgggg	4080
gaggggggag	ccagaatgag	gcgogcggt	aaccgaagtt	cctatacttt	ctagagataa	4140
ggaacttcgg	aataggaact	tcttaggtca	attctaccgg	gtaggggagg	cgcttttccc	4200
aaggcagtc	ggagcatgcg	cttttagcag	cccgctgggc	acttggcgct	acacaagtgg	4260
cctctggcct	cgcacacatt	ccacatccac	cggtaggcgc	caaccggctc	cgttcttttg	4320
tggccccctt	gcgccacctt	ctactcctcc	ctagtccag	aagttccccc	ccgc'cccgca	4380
gctcgcgctcg	tgaggagcgt	gacaaatgga	agtagcacgt	ctcactagtc	tcgtgcagat	4440
ggacagcacc	gctgagcaat	ggaagcgggt	aggcctttgg	ggcagcggcc	aatagcagct	4500
ttgctccttc	gctttctggg	ctcagaggct	gggaaggggg	gggtccgggg	gcgggctcag	4560
gggcccggctc	aggggcgggg	cgggcgccc	aaggctcctc	ggaggcccgg	cattctgcac	4620
gcttcaaaa	cgcacgtctg	ccgcgctgtt	ctcctcttcc	tcattctcgg	gcctttcgac	4680
ctgcagccaa	tatgggatcg	gccattgaac	aagatggatt	gcacgcaggt	tctccggccg	4740
cttgggtgga	gaggctattc	ggctatgact	gggcacaaca	gacaatcggc	tgctctgatg	4800
ccgcgctgtt	ccggctgtca	gcgcaggggc	gcccgggtct	ttttgtcaag	accgacctgt	4860
ccggtgccc	gaatgaactg	caggacgagg	cagcgcggct	atcgtggctg	gccacgacgg	4920
gcgttccttg	cgcagctgtg	ctcgacgttg	tcaactgaag	gggaaggggac	tggtctgtat	4980
tgggcgaagt	gccggggcag	gatctcctgt	catctcacct	tgcctcctgc	gagaaagtat	5040
ccattcatggc	tgatgcaatg	cggcggtgct	atacgcctga	tccggctacc	tgcccatctg	5100
accaccaage	gaaacatcgc	atcgagcgag	cacgtactcg	gatggaagcc	ggctctgtcg	5160
atcaggatga	tctggacgaa	gagcatcagg	ggctcgcgc	agccgaactg	ttcgccaggc	5220
tcaaggcgcg	catgcccagc	ggcgaggatc	tcgtcgtgac	ccatggcgat	gcctgcttgc	5280
cgaatatcat	ggtggaaaa	ggccgctttt	ctggattcat	cgactgtggc	cggctgggtg	5340
tggcgggaccg	ctatcaggac	atagcgttgg	ctaccgtga	tattgtctgaa	gagcttggcg	5400
gagaatgggc	tgaccgcttc	ctcgtgcttt	acggtatcgc	cgctcccgat	tcgcagcgca	5460
tcgcccctcta	tcgcccctctt	gacgagttct	tctgagggga	tcgatccgct	gtaagtctgo	5520
agaaattgat	gatctattaa	acaataaaga	tgtccactaa	aatggaagtc	tttccctgtca	5580
tactttgtta	agaaggggtga	gaacagagta	cctacatttt	gaatggaagg	attggagcta	5640

19

cgggggtggg	ggtgggggtgg	gattagataa	atgcctgttc	tttactgaag	gctcttttact	5700
attgctttat	gataatgttt	catagttgga	tatcataatt	taaacaagca	aaaccaaatt	5760
aagggccagc	tcattcctcc	cactcatgat	ctatagatct	atagatctct	cgtgggatca	5820
ttgtttttct	cttgattccc	actttgtggt	tctaagtact	gtggtttcca	aatgtgtcag	5880
tttcatagcc	tgaagaacga	gacagcagc	ctctgttcca	catacacttc	attctcagta	5940
ttgttttgcc	aagttctaat	tccatcagaa	gctgactcta	gatcccgccg	egaagttcct	6000
atactttcta	gagaatagga	acttcggaat	aggaacttc			6039

<210> 7

<211> 14411

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: targeting
vector for Rosa26 locus with a CAGGS-creER insert

<400> 7

tttgagttag	ctgataccgc	tgcgcgcagc	cgaacgaccg	agcgcagcga	gtcagttagc	60
gaggaagcgg	aagagcgcgc	aatacgcaca	ccgcctctcc	ccgcgcgttg	gccgattcat	120
taatgcagct	ggcagcagag	gtttcccgac	tggaaaagcgg	gcagttagcg	caacgcaatt	180
aatgtgagtt	agctcactca	ttaggcaccc	caggctttac	actttatgct	tccggctcgt	240
atgtttgtgt	gaattgtgag	cggataacaa	tttcacacag	gaaacagcta	tgaccatgat	300
tacgccaagc	gcgcaattaa	ccctcactaa	agggaaacaaa	agctgtcgag	atctagatat	360
cgatggccat	agagttacgc	tagggataac	agggtaatat	agccgcggca	ggccctccga	420
gcgtgggtgga	gccgttctgt	gagacagccg	ggtacgagtc	gtgacgctgg	aaggggcaag	480
cgggtgggtgg	gcaggaatgc	ggtccgcctc	gcagcaaccg	gagggggagg	gagaagggag	540
cggaaaagtc	tccacccggc	gcggccatgg	ctcggggggg	gggggggcagc	ggaggagcgc	600
ttccggccga	cgtctcgtcg	ctgattggct	tcttttcttc	ccgcctgtgt	tgaaaacaca	660
aatggcgtgt	tttggttggc	gtaaggcgcc	tgctcagttaa	cgccagccgg	agtgcgcagc	720
cgccggcagc	ctcgtctctg	ccactgggtg	gggcgggagg	taggtggggt	gaggcgagct	780
ggacgtgcgg	gcgcgggtcg	cctctggcgg	ggcgggggag	gggagggagg	gtcagcgaaa	840
gtagctcgcg	cgcgagcggc	cgcccaccct	ccctcttctc	tgggggagtc	gttttaccgc	900
ccgcggcccg	ggcctcgtcg	tctgattggc	tctcggggcc	cagaaaactg	gcccttgcca	960
ttggctcgtg	ttcgtgcaag	ttgagtcctt	ccgcggccca	gcggggggcg	cgaggaggcg	1020
ctcccaggtt	ccggccctcc	cctcggcccc	gcgcgcgaga	gtctggccgc	gcgccctgc	1080
gcaacgtggc	aggaagcgcg	cgtcgggggc	ggggacgggc	agtagggctg	agcggctgcg	1140
gggcgggtgc	aagcacgttt	cgcacttgag	ttgcctcaag	aggggcgtgc	tgagccagag	1200
ctccatcgcg	cactccgggg	agtggaggga	aggagcgagg	gctcagttgg	gctgttttgg	1260
agggcaggaag	cacttgctct	cccaaagtgc	ctctgagtcg	ttatcagtaa	gggagctgca	1320
gtggagtagg	cgggggagaag	gccgcaccct	tctccggagg	ggggagggga	gtgttgcaat	1380
acctttcttg	gagttctctg	ctgcctcctg	gcttctgagg	accgccttgg	gcctgggaga	1440
atcccttccc	cctcttccct	cgtgatctgc	aactccagtc	tttctaggta	accgatatcc	1500
ctgcaggttt	tgcacattga	ttattgacta	gttattaata	gtaatcaatt	acggggtcat	1560
tagttcatag	cccataatag	gagttccgag	ttacataact	tacggtaaat	ggcccgccctg	1620
gctgaccgcc	caacgacccc	cgccatttga	cgtcaataat	gacgtatgtt	cccatagtaa	1680
cgccaatagg	gactttccat	tgacgtcaat	gggtgggacta	tttacggtaa	actgcccact	1740
tggcagtaga	tcaagtgtat	catatgccaa	gtacgcccc	tattgacgtc	aatgacggta	1800
aatggcccg	ctggcattat	gcccagtaga	tgaccttatg	ggactttcct	acttggcagt	1860
acatctacgt	attagtcate	gctattacca	tgggtcgagg	tgagccccc	gttctgertc	1920
actctcccca	tctccccccc	ctccccacc	ccaattttgt	atttatttat	tttttaatta	1980
ttttgtgcag	cgatgggggg	gggggggggg	ggggcgcgcg	ccaggcgggg	cggggggggg	2040
cgagggggcg	ggcggggcga	ggcgagagag	tgcgggcgga	gccaatcaga	gcggcgcgct	2100
ccgaaagttt	ccttttatgg	cgaggcgggc	gcggcgggcg	ccctataaaa	agcgaagcgc	2160
gcggcgggcg	ggagtcgctg	cgttgccctc	gccccgtgcc	ccgctccgcg	ccgcctcgcg	2220
ccgcggccgc	cggtctctgc	tgaccgcggt	actcccacag	gtgagcgggc	gggacggccc	2280
ttctcctccg	ggctgttaatt	agcgtttggt	tttaatgacgg	ctcgcttctt	ttctgtggct	2340
gcgtgaaagc	cttaaagggg	tccgggaggg	ccctttgtgc	ggggggggagc	ggctcggggg	2400
gtgcgtgcgt	gtgtgtgtgc	gtggggagcg	ccgcgtgcgg	ccgcgcctgc	ccggcggtgc	2460
tgagcgctgc	gggcgcggcg	cggggcttctg	tgcgctccgc	gctgcgcgga	ggggagcgcg	2520

gcccggggggcg	gtgccccgcg	gtgccccggg	gctgccccggg	gaacaaaggc	tgcgtgccccg	2580
gtgtgtgtgt	gggggggtga	gcaggggggtg	tggggcgccg	ggtcggggctg	taaccccccc	2640
ctgcaccccc	ctccccgagt	tgctgagcac	ggccccggctt	cggtgccccg	gctccgtgcg	2700
gggcggtggcg	cgggggtcgc	cgtgccccggc	gggggggtggc	ggcaggtggg	ggtgccccggc	2760
ggggcgggggc	cgccctcgggc	cgggggagggc	tcggggggagg	ggcgccccgg	ccccggagcg	2820
cggcgggctg	tcgaggcgcg	gcgagccgcga	gccattgccc	tttatggtaa	tcgtgcgaga	2880
ggggcgaggg	acttcctttg	tcocaaatct	ggcgagcccg	aaatctggga	ggcgcccccg	2940
cacccccctct	agcgggcgcg	ggcgaagcgg	tcggggcgccg	gcaggaagga	aatggcgggg	3000
gagggccttc	gtgctgcgcc	gcgcggccgt	ccccttctcc	atctccagcc	tcggggctgc	3060
cgcaggggga	cggctgcctt	cggggggggac	ggggcgagggc	gggggttcggc	ttctggcgctg	3120
tgaccggcgcg	ctctagaagc	gttgggggtga	gtactccctc	tcacaaagcg	gcacgacctc	3180
tgcgctaaga	ttgtcagttt	ccaaaaacga	ggaggatttg	atattcacot	ggccccgggt	3240
gatgcctttg	agggtggccg	cgtccatctg	gtcagaaaag	acaatctttt	tggtgtcaag	3300
cttgaggtgt	ggcaggcctg	agatctggcc	atacacttga	gtgacattga	catccacttt	3360
gcctttctct	ccacaggtgt	ccactcccag	ggcgccctcc	ggagcgatcg	ccgggtccgcc	3420
taggctagcc	ggccgcgtcg	accatgtcca	attractgac	cgtacaccaa	aatttgcttg	3480
cattaccggc	cgatgcacac	agtgtgagg	ttcgcaagaa	cctgatggac	atggtcaggg	3540
atcgccaggc	gttttctgag	catacctgga	aaatgcttct	gtccggtttg	cgggtcggtgg	3600
cggcatggtg	caagttgaat	aaccggaaat	gggttcccg	agaacctgaa	gatgttcgcg	3660
attatcttct	atctcttcag	gcgcgcgggtc	tggcagtaaa	aactatccag	caacatttgg	3720
gccagctaaa	catgcttcat	cgtcgggtccg	ggctgcccag	accaagtga	agcaatgctg	3780
tttctactgt	tatgcggcg	atccgaaaag	aaaacgttga	tgccgggtgaa	cgtgcaaaac	3840
aggctctagg	gttcgaacgc	actgatttcc	accaggttcg	ttcactcatg	gaaaatagcg	3900
atcgctgcca	ggatatacgt	aatctggcat	ttctggggat	tgcttataac	acctgtttac	3960
gtatagccga	aattggcagg	atcagggtta	aagatatctc	acgtactgac	ggtgggagaa	4020
tgtaatacca	tattggcaga	acgaaaacgc	tgggttagcac	cgcaggtgta	gagaaggcac	4080
ttagcctggg	ggtaactaaa	ctggtcagag	gatggatttc	cgtctctggt	gtagctgatg	4140
atccgaataa	ctacctgttt	tgccgggtca	gaaataattg	tggtgcccgc	ccatctgcca	4200
ccagccagct	atcaactcgc	gcccctggaa	ggatttttga	agcaactcat	cgtatgattt	4260
acggcgctaa	ggatgactct	ggtcagagat	acctggccctg	gtctggacac	agtgcccgctg	4320
tcggagccgc	gcgagatatg	gcccgcgctg	gagtttcaat	accggagatc	atgcaagctg	4380
gtggctggac	caatgtaaat	attgtcatga	actatatccg	taacctggat	agtgaacacg	4440
gggcaatggt	gcgcctgctg	gaagatggcg	attctgctgg	agacatgaga	gctgccaacc	4500
tttggccaag	cccgctcatg	atcaaacgct	atcagaagaa	cagcctggcc	ttgtccctga	4560
cggccgacca	gatggtcagt	gccttgtttg	atgctgagcc	ooccatactc	tattccgagt	4620
atgatcctac	cagacccttc	agtgaagctt	cgatgatggg	cttactgacc	aacctggcag	4680
acagggagct	ggttcacatg	atcaactggg	cgaagagggt	gccagggctt	gtggatttga	4740
ccctccatga	tcagggtccac	cttctagaat	gtgcctggct	agagatcctg	atgattgggtc	4800
tcgtctggcg	ctccatggag	cacccagtga	agctactgtt	tgctcctaac	ttgctcttgg	4860
acaggaacca	ggggaataat	gtagagggtga	tggtggagat	cttcgacatg	ctgctggcta	4920
catcatctcg	gttcgcgcag	atgaatctgc	agggagagga	gtttgtgtgc	ctcaaatcta	4980
ttattttgct	taattctgga	gtgtacacat	ttctgtccag	cacctggaag	tctctggaag	5040
agaaggacca	tatccaccga	gtcctggaca	agatcacaga	cactttgatc	cactgatgg	5100
ccaaggcagg	cctgaacctg	cagcagcagc	accagcggtc	ggcccagctc	ctcctcatcc	5160
tctcccadat	caggcacatg	agtaacaaag	cgatggagca	tcgttacagc	atgaagtga	5220
agaacgtggt	gcccccttat	gacctgctgc	tggaggcgcc	cgacgcccac	cgcctacatg	5280
cgcgccactag	cbgtggaggg	gcacccgtgg	aggagacgga	ccaaagccac	ttggccactg	5340
cgggctctac	ttcatcgcat	tccttgcaaa	agtattacat	cacgggggag	gcagagggtc	5400
tcctgcccac	agtctgagcg	gccgaccggg	tcgagatcca	ggcgcggtat	aataaaagat	5460
cattatcttc	aatagatctg	tggtgtgggt	ttctgtgtgc	cttgggggag	ggggaggcca	5520
gaatgaggcg	cggccaaggg	ggaggggggag	gccgaatga	ccttggggga	gggggaggcc	5580
agaatgacct	tgggggaggg	ggaggccaga	atgaggcgcg	ccggtaaccg	aagtctccat	5640
actttctaga	gaataggaac	ttcggaatag	gaacttctta	ggtcaattct	accgggtagg	5700
ggaggcgctt	ttcccaaggc	agtctggagc	atgcgcttta	gcagccccgc	tgggcactctg	5760
gcgctacaga	agtggcctct	ggcctcgcac	acattccaca	tcacacggta	ggcgccaacc	5820
ggctccgctc	cttgggtggc	ccttcgcgcc	acctccctac	cctcccttag	tcagggaagt	5880
cccccccgcc	ccgcagctcg	cgtcgtgcag	gacgtgacaa	atggaaagtag	cacgtctcac	5940
tagtctcgrg	cagatggaca	gcaccgctga	gcaatggaa	cgggtaggcc	tttggggcag	6000
cggccaatag	cagctttgct	ccttcgcttt	ctgggctcag	aggctgggaa	gggggtgggtc	6060
cggggggcg	ctcagggggc	ggctcagggg	cggggcgggc	gcccgagggt	cctccggagg	6120
cccggcattc	tgacgccttc	aaaagcgcac	gtctgcccgc	ctgttctcct	cttccctcatc	6180

tccgggcctt	tgcacctgca	gccaatatgg	gatcggccat	tgaacaagat	ggattgcacg	6240
cagggttctcc	ggcgcgttgg	gtggagagggc	tattcggcta	tgactgggca	caacagacaa	6300
tccgctgctc	tgatgcgcgc	gtgttccggc	tgtcagcgca	ggggcgcccg	gttctttttg	6360
tcaagaccga	cctgtccggg	gccctgaatg	aactgcagga	cgaggcgagcg	cggtatccgt	6420
ggctggccac	gacgggcggt	ccttgccgag	ctgtgctoga	cggtgtcact	gaagcgggaa	6480
gggactggct	gctattgggc	gaagtgcggg	ggcaggatct	cctgtcatct	caecttgctc	6540
ctgccgagaa	agtatccatc	atggctgatg	caatgcggcg	gctgcatacg	cttgatccgg	6600
ctacctgccc	attcgaccac	caagcgaaac	atcgcatcga	gcgagcacgt	actcggatgg	6660
aagccgggtc	tgtcgatcag	gatgatctgg	acgaagagca	tcaggggctc	gcgccagccg	6720
aactgttcgc	cagggtcaag	gcgcgcacgc	ccgacgggca	ggatctcgtc	gtgacccatg	6780
gcatgacctg	cttgcgaat	atcatgggtg	aaaatggocg	ctttctctgga	ttcatcgact	6840
gtggccggct	gggtgtggcg	gaccgctatc	aggacatagc	gttggctacc	cgatgatattg	6900
ctgaagagct	tggcgccgaa	tgggctgacc	gcttccctgt	gctttacggg	atcgccgctc	6960
ccgattcgca	gcgcacgcgc	ttctatcgcc	ttcttgacga	gttcttctga	ggggatcgat	7020
cegetgtaag	tctgcagaaa	ttgatgatct	attaacaaat	aaagatgtcc	actaaaatgg	7080
aagtttttcc	tgctcatactt	tgttaagaaag	ggtgagaaca	gagtacctac	actttgaatg	7140
gaaggattgg	agctacgggg	gtgggggtgg	gggtgggatta	gataaatgcc	tgctctttac	7200
tgaaggctct	ttactattgc	tttatgataa	tgtttcatag	ttggatatca	taatttaaac	7260
aagcaaaaacc	aaattaaggg	ccagctcatt	cctccacttc	atgatctata	gatctataga	7320
tctctcgtgg	gatcattgtt	ttctctctga	ttccactttt	gtggttctaa	gtactgtggg	7380
ttccaaatgt	gtcagtttca	tagcctgaag	aacgagatca	gcagcctctg	ttccacatac	7440
acttcaactct	cagtattgtt	ttgccaaagt	ctaattccat	cagaagctga	ctctagatcc	7500
cgcgccgaag	ttcctatact	ttctagagaa	taggaacttc	ggaataggaa	cttcaagctt	7560
aagcgctaga	agatgggctg	gagtcctctg	ggcaggetta	aaggctaacc	tggtgtgtgg	7620
gcgttgtctc	gcaggggaat	tgaacaggtg	taaaatttga	gggacaagac	ttcccacaga	7680
ttttcgggtt	tgtcgggaag	tttttcaata	ggggcaataa	aggaaaatgg	gaggataggt	7740
agtcacctgg	gggtttatgc	agcaaaaacta	cagggttatta	ttgcttgtga	tcgcctcggg	7800
agtattttcc	atcgaggtag	attaaagaca	tgtctacccg	agttttatac	tctcctgctt	7860
gagatccctta	ctacagtatg	aaattacagt	gtcgcgagtc	agactatgta	agcagaatct	7920
taatcarttt	taaagagccc	agtacttcat	atccatttct	ccgctccttt	ctgcagcctt	7980
atcaaaaagg	atttttagaac	actcatttta	gccccatttt	catttattat	actggccttat	8040
ccaaccctta	gacagagcat	tggcattttc	ccttctctga	tcttagaagt	ctgatgactc	8100
atgaaaccag	acagatttag	tacatacacc	acaaatcgag	gctgtagctg	gggcctcaac	8160
actgcagttc	ttttataact	ccttagtaca	ctttttgttg	atcctttgct	ttgatccctta	8220
attttcagtg	tctatcacct	ctcccgtcag	tgggtgttcca	catttggggc	tattctcagt	8280
ccagggagtt	ttacaacaat	agatgtattg	agaatocaac	ctaaagctta	actttccact	8340
cccatgaatg	cctctctctc	ttttctccat	ttataaaactg	agctatttaac	cattaatggg	8400
tccaggttga	tgtctctctc	ccatattacc	tgatgtatct	tacataattgc	caggctgata	8460
ttttaagaca	ttaaaaggta	tatttccatta	ttgagccaca	tggtattgat	tactgtctac	8520
taaaaattttg	tcattgtaca	catctgtaaa	agggtggttcc	ttttgggaatg	caaagttcag	8580
gtgtttgttg	tctttctctga	cctaagggtct	tgtgagcttg	tattttttct	attttaagcag	8640
tgtttctctc	tggactgggt	tgactcatgg	cattctacac	gttattgctg	gtctaaatgt	8700
gattttgcca	agcttcttca	ggacctataa	ttttgtctga	cttgtagcca	aacacaagta	8760
aaatgattaa	gcaacaaatg	tattttgtgaa	gcttgggtttt	taggtttgttg	tggtgtgtgt	8820
gcttgtgtctc	tataataata	ctatccaggg	gctggagagg	tggtctggag	ttcaagagca	8880
cagactgctc	ttccagaagt	cctgagttca	attcccagca	accacatggg	ggctcacaac	8940
catctgtaat	gggatctgat	gccctcttct	ggtgtgtctg	aagaccacaa	gtgtattcac	9000
attaaataaaa	taaatcctcc	ttcttcttct	ttcttttttt	tttaaagaga	atactgtctc	9060
cagtagaattt	tactgaagta	atgaaataact	ttgtgtcttg	tocaatatgg	tagccaataa	9120
tcaaattact	ctttaagcac	tggaaatggt	accaagggaac	taatttttat	ctgaagtgtg	9180
actgttgaca	gaggagccat	aactgcagac	ttgtgggata	cagaagacca	atgcagactt	9240
taatgtcttt	tctcttacac	taagcaataa	agaaaataaaa	attgaacttc	tagtatccta	9300
tttgtttaaa	ctgctagctt	tacttaactt	ttgtgcttca	tctatacaaaa	gctgaaagct	9360
aagtctgcag	ccattactaa	acatgaaagc	aagtaatgat	aatttttggat	ttcaaaaatg	9420
tagggccaga	gttttagccag	coagtggctg	tgcttgccct	tatgccttta	atcccagcac	9480
tctggaggca	gagacaggca	gatctctgag	tttgagccca	gcctgggtcta	cacatcaagt	9540
tctatctagg	atagccagga	atacacacag	aaacctgtt	ggggagggggg	gctctgagat	9600
ttcataaaaat	tataattgaa	gcattcccta	atgagccact	atggatgtgg	ctaaatccgt	9660
ctacctttct	gatgagattt	gggtattatt	ttttctgtct	ctgctgttgg	ttgggtcttt	9720
tgacactgtg	ggctctcttt	aaagcctcct	tcctgccatg	tggctctctg	tttgctacta	9780
acttcccatg	gcttaaatgg	catggctttt	tgccttctaa	gggcagctgc	tgagatttgc	9840

agcctgattt	ccaggggtggg	gttgggaaat	ctttcaaaca	ctaaaattgt	cctttaattt	9900
tttttttaaa	aaatgggtta	tataataaac	ctcatataat	agttatgagg	agtgaagggtg	9960
actaatatta	aatgagtcce	tccctataaa	aagagctatt	aaggcttttt	gtcttatact	10020
taactttttt	tttaaatgtg	gtatctttag	aaccaagggt	cttagagttt	tagtatacac	10080
aaactggtgc	atcgcttaat	cagattttct	agtttcaaata	ccagagaato	caaatcttc	10140
acagccaaag	tcaaatataag	aattttotgac	ttttaatgtt	aatttgctta	ctgtgaatat	10200
aaaaatgata	gcttttctctg	agggcagggtc	tcactatgta	tctctgcctg	atctgcaaca	10260
agatatgtag	actaaagttc	tgcctgcttt	tgtctcctga	atactaagggt	taaaatgtag	10320
taatactttt	ggaaacttgca	ggtcagattc	ttttataggg	gacacactaa	gggagcttgg	10380
gtgatagttg	gtaaaatgtg	tttcaagtga	tgaaaacttg	aattattatc	accgcaacct	10440
acttttttaa	aaaaaaagcc	agggcctgta	gagcattgct	aagggatccc	taggacttgc	10500
tgagcacaca	agagtagtta	cttggcaggc	tctgggtgag	agcatattto	aaaaaacaaag	10560
gcagacaacc	aagaaactac	agttaagggt	acctgtcttt	aaaccatcrg	catatacaca	10620
gggatattaa	aatattccaa	ataatatttc	attcaagttt	tcccccatca	aattgggaca	10680
tggatttctc	cggtgaatag	gcagagttgg	aaactaaaca	aatgttgggt	ttgtgatttg	10740
tgaattgtt	ttcaagtgat	agttaaagcc	catgagatac	agaacaaaagc	tgctatttctg	10800
aggtctcttg	gtttatactc	agaagcactc	ctttgggttt	ccctgcacta	tcctgatcat	10860
gtcttagggc	taccttaggc	tgattgttgt	tcaaataaac	ttaagtttcc	tgctaggtga	10920
tgtcatatga	tttcatatat	caaggcaaaa	catgttatat	atgttaaaca	tttgactta	10980
atgtgaaagt	taggtctttg	tgggtttgat	ttttaatttt	caaaacctga	gctaaataag	11040
tcattttttac	atgtcttaca	tttgggtggaa	ttgtataaatt	gtgggtttgca	ggcaagactc	11100
tctgaacctag	taaccctacc	tatagagcac	tttgctgggt	cacaagtcta	ggagtcaagc	11160
atttcacctt	gaagttgaga	cgttttggta	gtgtatacta	gtttatatgt	tggaggacat	11220
gtttatccag	aagatattca	ggactatttt	tgactgggct	aagggaattga	ttctgatttag	11280
cactgttagt	gagcattgag	tggccttttag	gcttgaattg	gagtcacttg	tatatctcaa	11340
ataatgctgg	ccttttttaa	aaagcccttg	ttctctatca	ccctgttttc	tacataattt	11400
ttgttcaaag	aaatacttgt	ttggatctcc	ttttgacaac	aatagcatgt	tttcaagcca	11460
tatttttttt	cctttttttt	tttttttttg	gttttctcag	acagggtttc	tctgtatagc	11520
cctggctgtc	ctggaactca	ccttgttagc	caggctggcc	togaactcag	aaatccgctt	11580
gctctgctct	cctgagtgcc	gggattaaag	gcgtgcacca	ccacgcctgg	ctaagtctga	11640
tattttgtta	tataactata	accaatacta	actccactgg	gtggattttt	aattcagtc	11700
gtagtcttaa	gtgggtcttta	ttggcccttc	atrraaaatct	actgttcact	ctaacagagg	11760
ctgttggtac	tagtggcact	taagcaactt	cctacggata	tactagcaga	ttaagggtca	11820
gggatatgaa	ctagtctagc	gttttgtata	cctaccagct	ttatactacc	ttgttctgat	11880
agaaatattt	caggacatct	agcaccatct	tgcctctata	gtgagtcgta	ttacaattca	11940
ctggcctgtc	ttttacaacg	tctgtactgg	gaaaaccttg	gcgttaacca	acttaatcgc	12000
cttgccagcac	atcccccttt	cgccagctgg	cgtaatatcg	aagaggcccg	caccgatcgc	12060
ccttcccaac	agttgcgcag	cctgaatggc	gaatgggacg	cgccctgtag	cgccgcatta	12120
agcgcggcgg	gtgtgggtgt	tacgcgcagc	gtgaccgcta	cacttgccag	cgccctagcg	12180
cccgcctcct	tcgctttcct	cccttccttt	ctgcgccagt	tcgcccggct	tccccgtcaa	12240
gctctaaatc	gggggctccc	tttagggttc	cgatttagtg	cctttacggca	cctcgacccc	12300
aaaaaacttg	attaggggtga	tgggtcacgt	agtggggcat	cgccctgata	gacgggtttt	12360
cgccctttga	cggtggagtc	caggttcttt	aatagtggtg	tcttggtcca	aactggaaca	12420
acactcaacc	ctatctcggt	ctattctttt	gattttataag	ggatttttgc	gatttcgggc	12480
tattgggttaa	aaaatgagct	gatttaacaa	aaatttaacg	ogaattttta	caaaatatta	12540
acgcttacaa	tttaggtggc	acttttcggg	gaaatgtgcg	cggaaccctt	atttggttat	12600
tttcttaaat	acattcaaat	atgtatccgc	tcattgagaca	ataaccctga	taaatgcttc	12660
aataaatattg	aaaaaggaag	agtatgagta	ttcaacattt	ccgtgtcgcc	cttattccct	12720
tttttgccgc	attttgccct	cctgtttttg	ctcaccacga	aacgctgggtg	aaagtaaaag	12780
atgctgaaga	tcagttgggt	gcacgagtg	gtracactga	actggatctc	aacagcggta	12840
agatccttga	gagttttcgc	cccgaagaao	gthttccaat	gatgagcact	tttraagttc	12900
tgctatctgg	cgcggtatta	tcccgtattg	acggccggca	agagcaactc	ggtcgcccga	12960
tacactatgc	tcagaaatgac	ttgggtgagt	actcaccagt	cacagaaaag	catcttacgg	13020
atggcatgac	agtaagagaa	ttatgcagtg	ctgccataac	catgagtgat	aacactgcgg	13080
ccaacttact	tctgacaacg	atcggaggac	cgaaggagct	aaccgctttt	ttgcacaaca	13140
tgggggatca	tgttaactcgc	cttgatcgtt	gggaaccgga	gctgaatgaa	gccataccaa	13200
ctgagcagcg	tgacaccacg	atgcctgtag	caactggcaac	aacgttgcgc	aaactactaa	13260
ctggcgaaot	acttactcta	gcttcccgcc	aaactgtaac	agactggatg	gaggcggata	13320
aagttgcagg	accacttctg	cgctcggccc	ttccggctgg	ctgggtttatt	gctgataaat	13380
ctggagccgg	tgagcgtggg	tctcgcggta	tcattgcagc	actggggcca	gatggtaagc	13440
cctcccgatat	cgtagttatc	tacacgacgg	ggagtcaggc	aactatggat	gaacgaaata	13500

23

gacagatcgc	tgagataggt	gcctcactga	ttaagcattg	gtaactgtca	gaccaagttt	13560
actcatatat	acttttagatt	gatttataaac	ttcatttttta	atttataaagg	atctagggtga	13620
agatccctttt	tgataatctc	atgaccataaa	tcccttaacg	tgagtttttcg	ttccactgag	13680
cgtcagaccc	cgtagaaaaag	atcaaaggat	cttcttgaga	tcctttttttt	ctgcgcgtaa	13740
tctgctgett	gcaaacataaa	aaaccaccgc	taccagcggg	ggcttggttg	ccggatcaag	13800
agctaccaac	tctttttccg	aaggtaactg	gcttcagcag	agcgcagata	ccaaatactg	13860
tccttctagt	gtagccgtag	ttaggccacc	acttcaagaa	ctctgtagca	ccgcctacat	13920
acctcgcctc	gctaatactg	ttaccagtgg	ctgctgccag	tgccgataag	tcgtgtctta	13980
ccgggttgga	ctcaagacga	tagttaccgg	ataaggcgca	gcggtcgggc	tgaacggggg	14040
gttcgtgcac	acagcccage	ttggagcgaa	cgacctacac	cgaactgaga	tacctacage	14100
gtgagctatg	agaaagcgcc	acgcttcccg	aagggagaaa	ggcggacagg	tatccggtaa	14160
gcggcagggg	cggaaacagga	gagcgcacga	gggagcttcc	agggggaaac	gcctgggtatc	14220
tttatagtc	tgctgggttt	cgccacctct	gacttgagcg	tcgatttttg	tgatgctcgt	14280
cagggggggc	gagcctatgg	aaaaacgcca	gcaacgcggc	ctttttacgg	ttcctggcct	14340
tttgctggcc	ttttgctcac	atgttctttc	ctgcgttatc	ccctgattct	gtggataacc	14400
gtattaccgc	c					14411

Claims

1. Method for generating transgenic eukaryotic cells having a modified Rosa26 locus which method comprises the following step
 - (a) introducing a functional DNA sequence into the Rosa26 locus of starting eukaryotic cells by homologous recombination with a targeting vector comprising said functional DNA sequence flanked by DNA sequence homologous to the Rosa26 locus.
2. The method of claim 1, wherein the eukaryotic cells
 - (i) are derived from a multi-cell organism including vertebrates invertebrates and plants, preferably are vertebrate cells, more preferably are derived from a mammal, including rodents such as mouse, rat, etc. or a fish such as zebrafish; and/or
 - (ii) are primary cells or immortalized cells; most preferably the cells are mammalian embryonic stem (ES) cells.
3. The method of claim 1 or 2, wherein
 - (i) the functional DNA sequence is a gene expression cassette comprising a gene of interest operatively linked to a promoter or is a DNA sequence which can be converted into such gene expression cassette; and/or
 - (ii) the gene of interest is selected from recombinases, reporter genes, receptors, signaling molecules, transcription factors, pharmaceutically active proteins and peptides, drug target candidates, disease causing gene products, toxins, etc.; and/or
 - (iii) the promoter is a ubiquitous or tissue specific promoter, either constitutive or inducible, preferably is a CAGGS, hCMV, PGK, FABP, Lck, CamKII, CD19, Keratin, Albumin, aP2, Insulin, MCK, MyHC, WAP, Col2A, Mx, tet or Trex promoter; and/or
 - (iv) the DNA sequences homologous to the Rosa26 locus are 0.2 to 20 kB, preferably 1 to 10 kB long; and/or

25

- (v) the functional DNA sequence or gene expression cassette further comprises one or more additional functional sequences including but not limited to marker genes; recombinase recognition sites, poly A signal, introns, etc.; and/or
- (vi) the targeting vector further comprises tags for protein detection, enhancers, selection markers, etc.

4. The method of claim 3, wherein the transgenic eukaryotic cells are derived from mouse, the DNA sequences homologous to the Rosa26 locus are derived from the 5' and 3' flanking arm of the mouse Rosa26 locus, preferably said homologous DNA sequences having the sequences shown in SEQ ID NO:4 and 5, respectively, and the promoter is a CAGGS-promoter, most preferably the targeting vector has the sequence shown in SEQ ID NO:7.
5. The method according to any one of claims 1 to 4, which further comprises one or more of the steps
 - (b) isolating the eukaryotic cells, preferably the ES cells having the desired functional DNA sequence integrated into the Rosa26 locus; and/or
 - (c) modifying the integrated functional DNA sequence and isolating ES cells having the desired modified functional DNA sequence.
6. A targeting vector as defined in claims 1 to 4.
7. A eukaryotic cell having a modified Rosa26 locus obtainable by the method of claims 1 to 5.
8. A method for preparing transgenic multi-cell organism having a modified Rosa26 locus which comprises utilizing the method as defined in claims 1 to 5.
9. The method of claim 8, wherein the transgenic multi-cell organism is a non-human mammal and said method comprises modifying an ES cell as defined in claims 1 to 5.

10. The method of claim 9 which further comprises one or more of the steps
(d) injecting ES cells obtained in steps (b) or (c) into blastocysts; and/or
(e) generating transgenic non-human animals carrying one or more functional genes of interest at the Rosa26 locus.
11. A transgenic multi-cell organism and a transgenic non-human mammal obtainable by the method of claim 8 and 9-10, respectively, and having an operatively functional gene expression cassette integrated into its Rosa26 locus.
12. Use of the eukaryotic cell of claim 7, the transgenic multi-cell organism of claim 11, or the transgenic non-human mammal of claim 11 for gene function studies, drug development, as disease model animals, etc.

Abstract

The invention provides a method for targeted transgenesis using the Rosa26 locus. Suitable nucleotide acid sequences and vectors for the targeted transgenesis are provided. The Rosa26 locus proved to be a suitable integration site allowing strong and predictable expression of inserted transgenes carrying exogenous promoters.

-1/3-

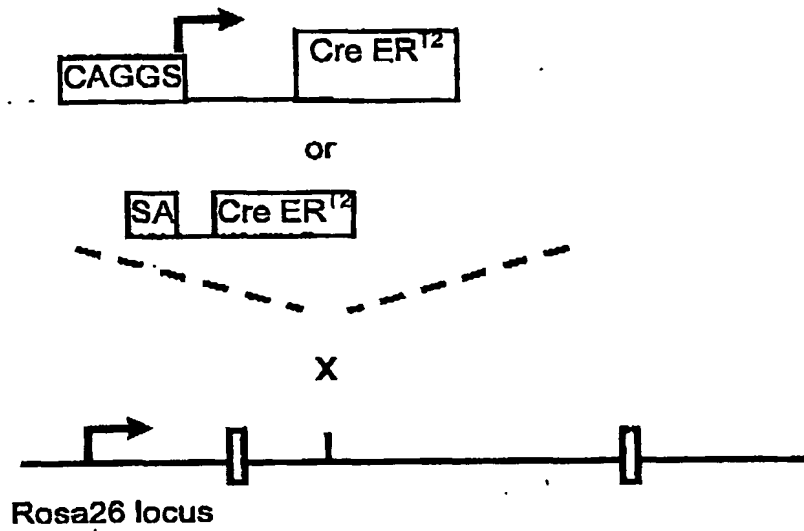
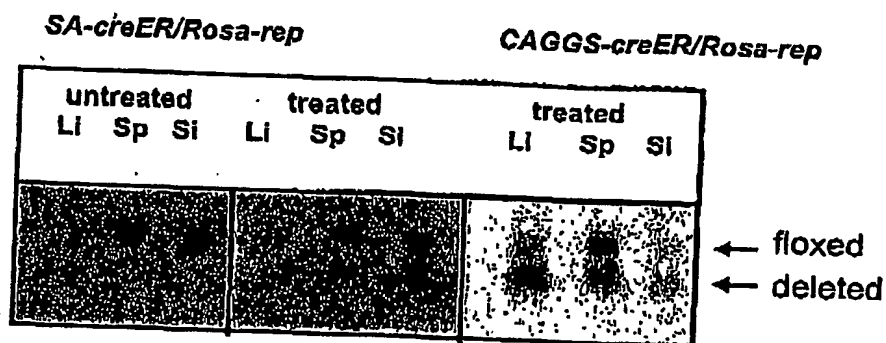
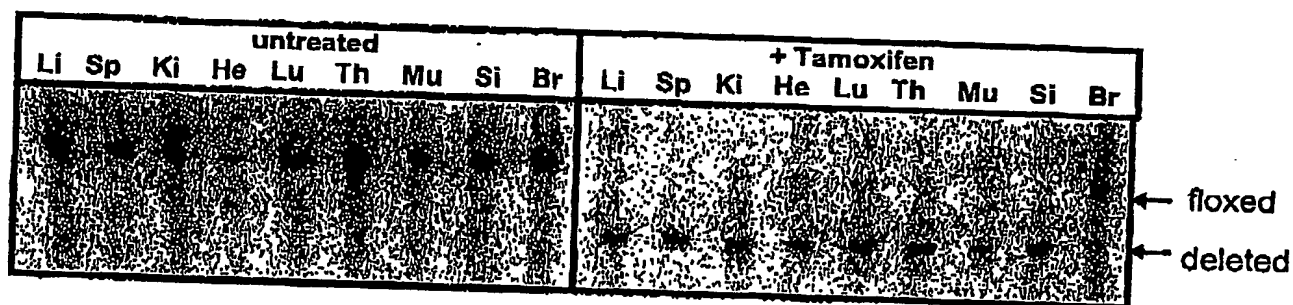


Fig.1

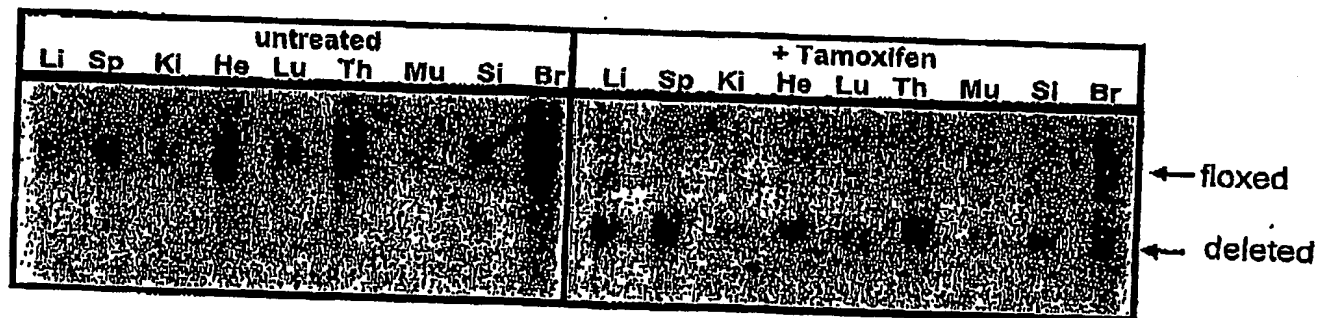
Fig.2



B

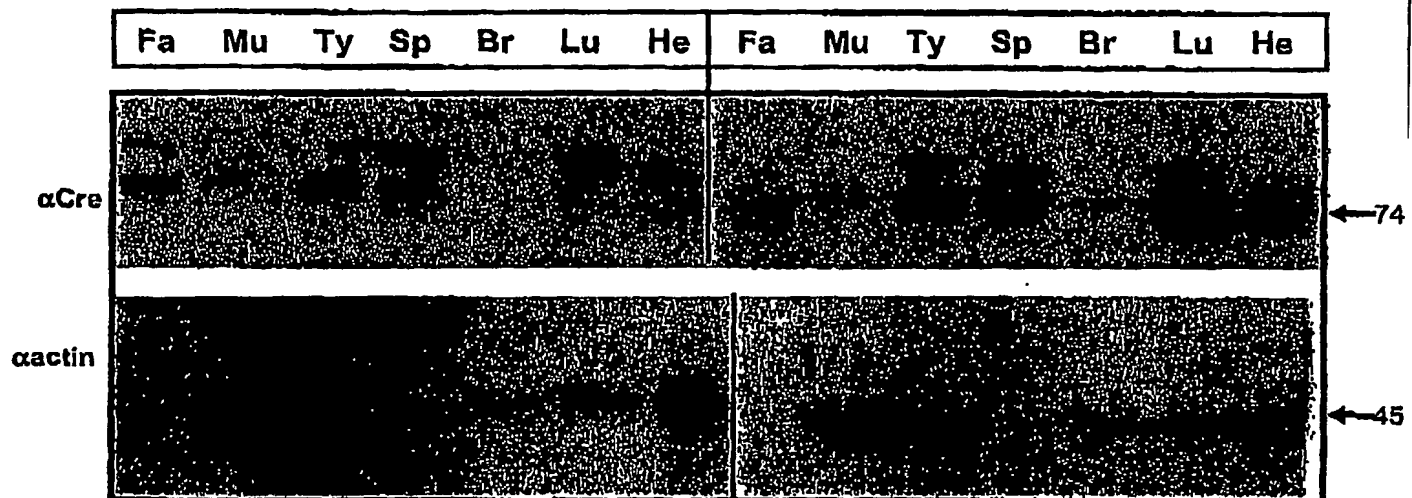


C



-3/3-

Fig.3



**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☒ **BLACK BORDERS**
- ☒ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☒ **FADED TEXT OR DRAWING**
- ☒ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **KEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☐ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.